Part 5- Session Papers for the EPA 23rd Annual National Conference on Managing Environmental Quality Systems

April 13-16, 2004   Tampa, Florida

Ambient Air I

Does Pulsed Sampling Provide Good Estimates of Atmospheric Concentrations? (Roger Carter and Debbie Lacroix, NOAA)

National Air Toxics Trends Stations Quality System (Dennis Mikel and Danny France, U.S. EPA)

Ambient Air II

Contributions of Tribal Environmental Agencies to Quality Assurance in Environmental Monitoring (Melinda Ronca-Battista, Northern Arizona University)

Multi-Site Evaluations of Candidate Methodologies for Determiningg Coarse particulate Matter (PMc) Concentrations (Robert Vanderpool, U.S. EPA and Andrew Johnson, Maine DEP)

Ambient Air III

Data Quality Objective Development for the Coarse Particualte Matter (PMc) Standard (Mike Papp, U.S. EPA)

New Quality Indicator Statistics for the Gaseous Criteria Pollutants (Basil Coutant, Battlelle)

Status and Changes in EPA Infrastructure for Bias Traceability to NIST (Mark Shanis, U.S. EPA)

Ambient Air IV

2003 EPA Audit of Protocol Cylinder Gases (Joe Elkins and B. Wright, U.S. EPA)

NPEP Through the Probe Demonstration (Greg Noah, U.S. EPA)

Does Pulsed Sampling Provide Good Estimates of Atmospheric Concentrations?

Debbie J. Lacroix, Roger G. Carter, NOAA-Air Resources Laboratory Field Research Division

For many years, the NOAA Air Resources Laboratory Field Research Division has used whole air bag samplers to sample atmospheric tracers. The samplers use small, battery operated diaphragm pumps to pump air into Tedlar bags for later analysis. To extend the sampling period for each bag and run the pumps within their performance range, the pumps are "pulsed", or turned on for a fraction of a second every few seconds. The assumption was made that the sample created by filling the bag with the pulses adequately represented the average concentration of the sampled species. Although this assumption should generally be good, its validity could be questioned close to the release point and in complex wind flows such as urban environments. To test the validity of this assumption, several experiments were conducted. A system of mass flow controllers was used to mix ultra pure air and sulfur hexafluoride (SF_6) as they flowed into a PVC pipe. The flow rates could be controlled to create a rapidly varying concentration of SF_6 within the pipe. Several "pulsed" samplers and continuous analyzers were allowed to sample simultaneously from the pipe and the resulting concentration measurements compared under various conditions. The results of the experiments indicate that the "pulsed" sampling method showed little difference from the true average concentrations in the pipe even under rapidly varying concentrations.

INTRODUCTION

Understanding how the atmosphere transports and disperses materials released into the air is essential to understanding and appropriately dealing with important problems such as air pollution and accidental toxic gas releases. Atmospheric tracer experiments play a key role in this process by providing basic information about transport and dispersion processes. An atmospheric tracer experiment consists of releasing a small amount of stable, non-toxic, easily detectable substance (known as a tracer) into the air. The air in the surrounding area is sampled and the concentration of the tracer is measured. By combining the tracer concentrations with meteorological information, scientists can develop and test theories and models of atmospheric dispersion. NOAA's Air Resources Laboratory Field Research Division (NOAA-ARLFRD) has sampled atmospheric tracers for many years with the use of bag samplers. These consist of 12 miniature diaphragm pumps and controlling electronics attached inside of a waxed cardboard box and are referred to as programmable integrating gas samplers (PIGS), or simply "samplers" (Figure 1). A smaller box containing 12 Tedlar bags is referred to as a "cartridge"(Figure 2). This cartridge is placed inside of the sampler and each bag is attached to a rubber tube which both supports the bag and provides an entrance and exit for the sampled air (Figure 3). A plastic clip located on the tube provides a means to seal the bag when sampling or analysis is not in progress. The lid is placed on the sampler and a bungee cord keeps the lid in place (Figure 4).
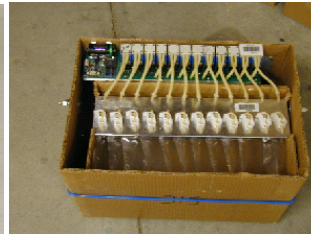
Figure 1. Sampler.

Figure 2. Cartridge.

Figure 3. Sampler with cartridge inside.

Figure 4. Sampler and cartridge with lid on and bungee hold down.

Typically, PIGS are placed at pre-selected sites prior to the start of an experiment and programmed to collect samples over the period of the experiment. The samplers are usually hung 10 feet (3 meters) above the ground on telephone poles or other structures (Figure 5). Beginning at the programmed start time, the controlling electronics sequentially fill the 12 bags with each bag being filled for the programmed sampling duration. After the experiment is over, the cartridges are removed and taken to the laboratory and analyzed. A new set of cartridges were sometimes loaded into the PIGS so sampling may continue while the analysis of the previous samples was taking place. By placing a relatively large number of PIGS on arcs or on a grid across the experimental area, a good footprint of the tracer plume can be determined for each sampling period.

Figure 5. Sampler being hung.

To prevent the bags from being overfilled, the pumps are not run continuously, but are "pulsed" for short bursts or "strokes" that last less than one second. The time between these strokes is calculated to provide a full, but not over inflated, bag at the end of the sampling duration and is typically between 4 and 16 seconds. This process assumes that the pump strokes will provide a sample that adequately reflects the average concentration during the sampling period. Typically, tracer concentrations don't exhibit large concentration changes in the few seconds between pump pulses and this assumption should be good. However, in some recent studies, where sampling was done relatively near the tracer release point, continuous analyzers like the one in Figure 6 showed some rapid, large concentration changes. (See Fig. 9.) In this case, there is a concern that the pulse method may not generate a good average concentration.

Figure 6. Continuous analyzer.

This is a essentially a discrete sampling problem. Much work has been done in this area over the years (Brigham) which centers around the well know sampling theorem. Basically, this says that if sampling is not fast enough (i.e. at the Nyquist frequency), the actual signal can't be reconstructed. Where there are large, rapid variations similar to those in Figure 9, the pulsed sampling may not be fast enough to reconstruct the entire signal, but the objective is to reconstruct only the average and not the entire signal. To check the effect of "under sampling" on the concentration average, three experiments were conducted using a setup which allowed concentrations to be varied rapidly.

EXPERIMENTAL SETUP

The experimental set up is shown in Figure 7. A six-foot (1.8 meter) long, 2-inch (5 cm) diameter PVC pipe was suspended horizontally. SF_6 calibration gas and ultra pure air were injected into one end of the pipe. The flows of the gases were set with the mass flow controllers as seen in the bottom of Figure 7. The other end of the pipe was vented to the outside of the building by means of a vinyl hose. Halfway down the pipe (3 feet from the injection end), nine tubing ports were installed. These were equally spaced around the circumference of the pipe. (See Figure 8) The samplers and continuous analyzers were attached to these sampling ports. For Test C, three additional ports were added 6.5 inches further down the tube, (i.e. 42.5 inches from the injection end of the pipe). Ports that were not being used during a given test were closed off by attaching a short piece of tubing and closing it with it clip.

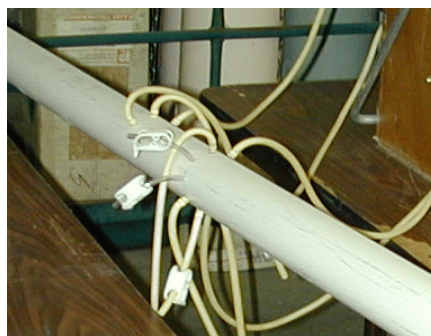**Figure 7. PVC pipe with tubing ports and flow meters attached.**

Figure 8. Closer view of tubing ports.

Three tests were designed to test the effect of the pulsed sampling method when sampling a poorly mixed atmosphere. Mass flow controllers were used to pulse and send puffs of the SF_6 calibration gas flowing into the pipe.

1. Since it is impossible to define a "typical" case of rapid concentration variation, Test A attempted to generate conditions that are worse than would be ever seen in nature. Thus, any problems observed in actual experiments should be less than those observed in this test. Instead of occasional rapid variations, the concentration was varied during the entire
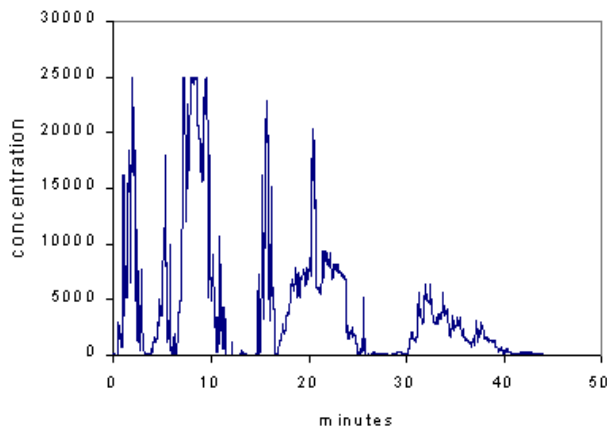
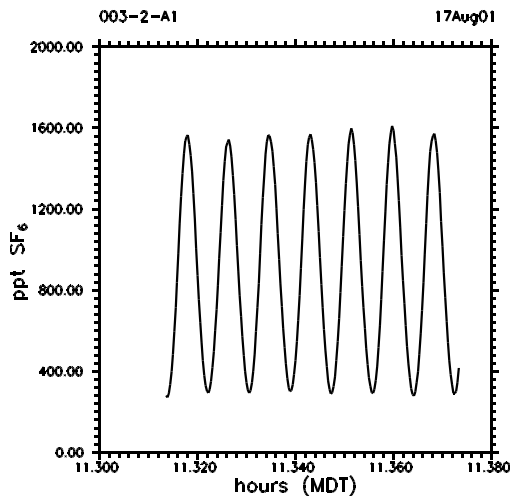**Figure 9. Example of continuous analyzer field measurement.**

Figure 10. Continuous analyzer output from Test A.

test. Also, the pump pulse period (30 sec.) exactly matched the period of concentration variation. Figure 10 shows the concentration as measured by a continuous analyzer (Figure 6). According to sampling theory, this should provide the maximum effect on the average concentrations. If a sampler happened to pulse at the concentration peaks, it would measure a much higher concentration than if it pulsed in the valley. An 18.6 ppbv SF_6 standard was cycled on and off and mixed with clean air to create the concentration variations. Two continuous analyzers (Figure 6) were used to verify that the gases didn't mix to a constant concentration in the pipe.

2.   Test B was set up exactly the same as test A with the exception that the sample pump pulses were so close together that the pumps ran continuously. Test B should provide the best possible average concentration. A comparison between tests A and B should show the effect on pulsed sampling at too slow of rate on average concentrations.

3.   Test C was an attempt to set up a sampling scenario that was a little more "real-world" than test A. It was a continuation of the pulsed sampling tests begun in Test A except a 19.9 ppmv SF_6 standard was diluted to about 24,000 pptv. Three extra ports were added downstream from the existing ports. The mass flow controllers were programmed to provide 1 second pulses with the time between pulses varying randomly from 0 to 10 seconds. This should provide an average concentration of about 4000 pptv. However, since the flow controllers do not open instantaneously, the actual average concentration was significantly lower than this. The pulse period was randomized to better approximate the more random processes of atmospheric mixing. The samplers were programmed to pulse their pumps every 5 seconds. Three of the samplers were moved to new ports about 6.5 inches down the tube. The distance of this move was calculated to provide approximately the same "flow time" difference as would occur between duplicate samplers under light wind conditions.

TEST RESULTS

The average concentration, standard deviation and relative standard deviation (RSD, the standard deviation divided by the average), for all three tests is shown in Table 1. The relative percent differences (RPDs, the difference divided by the average), of the average concentrations and RPD's of the RSD's are shown in Table 2. The RPD of the average concentration could only be calculated for Test A and B. Test C was run at a slightly different concentration and therefore could not be compared with the other two tests. The RPD's of the RSD's have therefore been calculated for all three tests in order to compare their deviations.

Test A, the "worst case scenario", resulted in the highest RSDs, with Test B, the "best case scenario", and Test C, the "real-world scenario" being fairly equivalent as seen in Table 1 below. This shows that the "worst case scenario" of sampling a poorly mixed atmosphere where the pumps pulse at the same frequency as the SF_6 puffs did in fact produce greater variability in the data. The precision was much better when the tracer pulses were more random as in Test C. Since real-world sampling should involve noticeable randomness, results from actual sampling should more closely approximate Test C.

Test C, the "real world scenario" mimicked Test B, the "best case scenario" extremely well with an RSD of 7% and 6% respectively as shown in Table 1. The RPD of the RSD between Tests B and C was 11%, much better than that between Tests A and B at 67% and Tests A and C at 57% as seen in Table 2. Test A, although extremely variable, still resulted in sampling a representative portion of the atmosphere as seen by the average concentration that was close to the result from Test B. The RPD of Test A and Test B was 8.5% which is well within the quality control limits of ±20% for the analysis. These results indicate that the pulsed sampling method does an adequate job of sampling the tracer over time even when the tracer is not well mixed in the atmosphere.

As seen in Table 1 below, the spread of the data (standard deviation) in Test A is approximately twice that of Test B. There was no indication in Test A of some higher cartridges and others that tended to be lower, creating a bias in the data, as was anticipated from the synchronized sampling. It doesn't appear that the worst case (Test A) for pulsed sampling was able to generate a statistically significant difference in the average concentration of all the bags in each cartridge. The only observable effect was an increase in variation in each bag causing a large amount of imprecision in this sampling method. These results indicate that there is no significant difference between the average pulsed samples and the continuous sampled average concentrations when the atmosphere is not well mixed, however, there is a significant difference in the variability between these two tests as seen by the variance. Because of the increased variations, individual samples may not yield good results. With an RSD of 16%, the 95% confidence limits would be ±32% and individual samples may be 32% off of actual concentrations. This is beyond the analysis quality control limits of ±20% by about 12%. Since Test A was designed to be significantly worse than actual conditions, it is not likely that significant effects would be seen in actual experiments.

Table 1. Results of Tests A, B and C

Test	Average Concentration (pptv)	Standard Deviation (pptv)	RSD (%)
A	979	155	16
B	1032	63	6
C	1422	96	7

Table 2. RPD of the average concentration and RPD of the average RSD between Tests A, B and C

Test	RPD of Average Concentration (%)	RPD of the RSDs (%)
A and B	8.5	67
A and C	NA	57
B and C	NA	11

SUMMARY

Tests A, B and C were designed to check the adequacy and accuracy of the pulsed sampling method in an atmosphere that is not well mixed. Test A was designed to generate "worst case scenario" data while Test B was designed to generate "best case scenario" data. Test C was designed to generate "real-world" data. Test A resulted in much more variable data, but an RPD of the average concentration result was within 10% of Test B. Test B and Test C resulted in similar data with an average RSD difference of only 1%. These tests can not be regarded as a complete characterization of the effects of pulsed sampling, but they do indicate that there are not large problems associated with the technique even in the case of rapid concentration fluctuations.

REFERENCES
Brigham, E. Oran, "The Fast Fourier Transform". Prentice-Hall, Inc, 1974, pp80-87.

National Air Toxics Trends Stations Quality System

Dennis K. Mikel, U.S. EPA
Office of Air Quality, Planning and Standards
Emission, Monitoring, and Analysis Division
Measurement, Methods and Technology Group
MD D243-02
Research Triangle Park, North Carolina, 27711
919-541-5511
mikel.dennisk@epa.gov

Abstract

In order to assess whether the population is at risk to Hazardous Air Pollutants (HAPs), the EPA is cooperation with the State/Local/Tribal (S/L/T) air pollution agencies are creating a long term national trend network. This paper will outline the Quality Assurance (QA) System that has been implemented for the National Air Toxics Trends Stations (NATTS) network In addition, EPA, in cooperation with the California Air Resources Board, has been sending out Proficiency Test (PT) samples to laboratories that had participated with the National Air Toxics Pilot Study, that was conducted in 2001. This paper will discuss the results of those PT samples and illustrate how that data will influence the future testing program for the NATTS.

Introduction

There are currently 188 HAPs, or Air Toxics (AT), regulated under the Clean Air Act (CAA) that have been associated with a wide variety of adverse health effects, including cancer, neurological effects, reproductive effects and developmental effects, as well as eco-system effects. These air toxics are emitted from multiple sources, including major stationary, area, and mobile sources, resulting in population exposure to these air toxics as they occur in the environment. While in some cases the public may be exposed to an individual HAP, more typically people experience exposures to multiple HAPs and from many sources. Exposures of concern result not only from the inhalation of these HAPs, but also, for some HAPs, from multi-pathway exposures to air emissions. For example, air emissions of mercury are deposited in water and people are exposed to mercury through their consumption of contaminated fish.

Our current Government Performance Results Act (GPRA) commitments specify that it is a goal to reduce air toxics emissions by 75% from 1993 levels to significantly reduce the risk to Americans of cancer and other serious adverse health effects caused by airborne toxics. EPA is developing new assessment tools and begin to address the risk associated with these emissions as required by the CAA and will be modifying that goal to one that focuses on risk reductions associated with exposure to air toxics Eventually, we have a long-term goal of eliminating unacceptable risks of cancer and other significant health problems from exposures to air toxics emissions and to substantially reduce or eliminate adverse effects on our natural environment.

History

In 2000, the EPA introduced the Air Toxics Concept Paper[1], which was a guideline that illustrated EPA's vision on air toxics monitoring in the future. One concept from the Concept Paper was to create a "Pilot" trends program, annual in its duration, which would help the S/L/T and EPA prepare for long term monitoring. The Pilot Program was implemented in 2001, with 10 stations in operation. After the successful implementation of the Pilot Program, Congress granted long term funding for a trends network. Planning for the NATTS began in 2002 with EPA soliciting requests to the States and Local air pollution agencies to work in cooperation in putting together a national trends network. The network has been set at 22 stations throughout the country. The NATTS began sampling January 1, 2004. In 2002-2003 EPA Office of Air Quality and Standards (OAQPS) began to create a Quality System that would enhance and allow quality professionals the ability to assess the uncertainty of data collected. Part of the Quality System is the development of guidance documents. To date, OAQPS has created the following guidance documents: Quality Management Plan, model Quality Assurance Project Plan[2], and a Technical Assistance Guidance[3] document. The Quality System is illustrated in Figure 1.
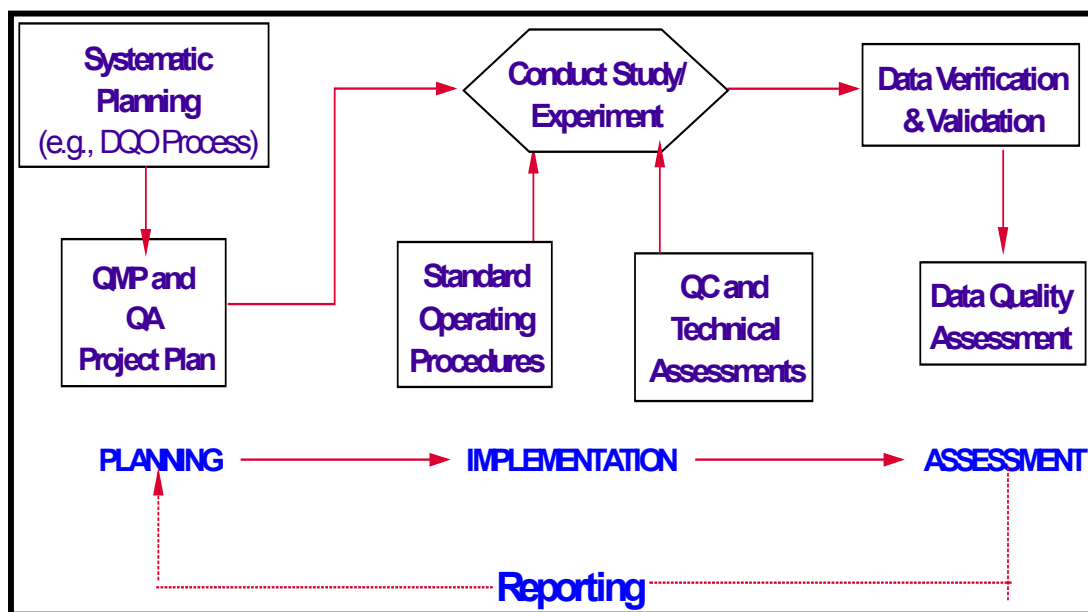

Figure1. NATTS Quality System

In 2001, OAQPS, in cooperation with the California Air Resources Board (CARB) and South Coast Air Quality Management District (SCAQMD) implemented a "Whole Air Sample" QA project that would allow EPA to assess the bias of the data that was collected during the Pilot Program. There were two types of samples collected: Particulate Matter -- 10 microns (PM_{10}) on quartz filters and ambient 3-hour integrated Volatile Organic Compounds (VOCs) samples collected inside of passivated stainless steel canisters. Both sets of samples were collected in Los Angeles, California.

The VOC samples were collected and distributed to 10 laboratories that were supporting the Pilot Program. The PM$_{10}$ filter samples were cut into strips and distributed to 5 laboratories. The findings of the 2001 Pilot Program inter-comparison study are documented elsewhere[4].

After the Pilot Program ended, it was decided to continue the VOC Whole Air Sample PT program. The initial 10 laboratories that had participated in the Pilot Program agreed to analyze blind PT samples that were prepared by CARB. This program has continued and data now exists for the 10 laboratories for 2001-2003.

Results

This section of the paper will discuss the results of the Whole Air PT Samples from 2001 through 2003. The author first analyzed the data to see which compounds were reported and how many laboratories reported those compounds. Below are the results of the 3 year comparison.
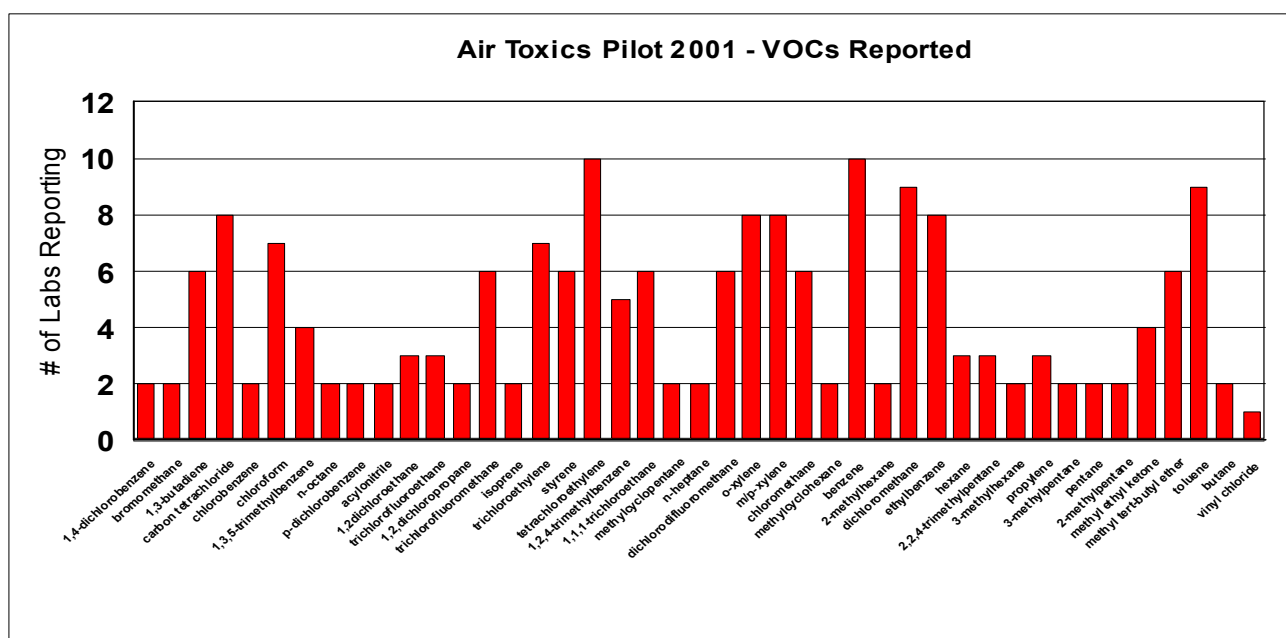

Figure 2. Distribution of Air Toxics Analyzed – 2001
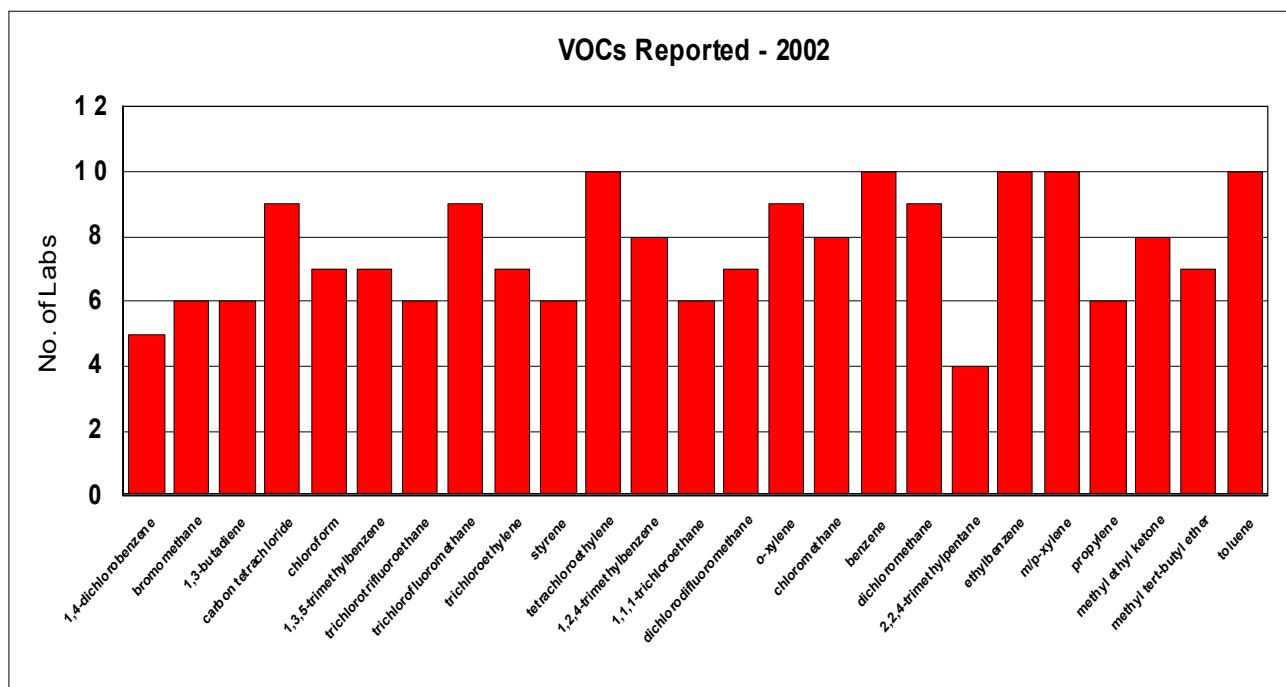
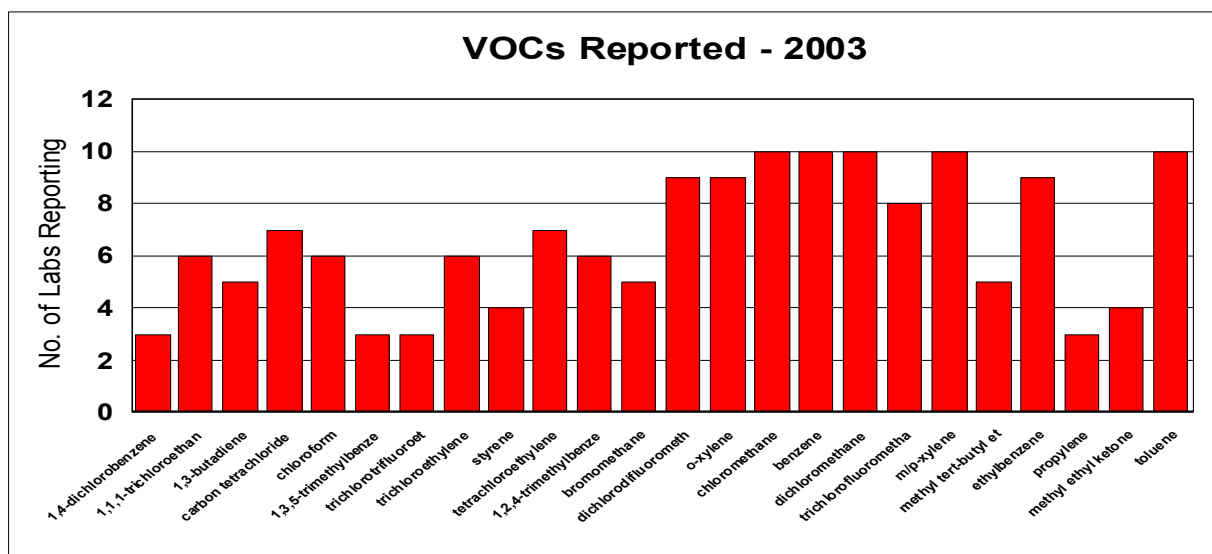**Figure 3. Distribution of Air Toxics Analyzed – 2002**

Figure 4. Air Toxics Analyzed - 2003

It can be seen from the analysis of the three years of data that in 2001, more than twice as many compounds were reported than in 2002 and 2003 also the overall means for all common compounds have decreased. The reason for this is unknown.

Compound	2001	2002	2003
Tetrachloroethylene (PERC)	10	10	7
Benzene	10	9	10
Ethyl-Benzene	8	10	7
m/p Xylene	8	10	8
Toluene	9	10	10
Carbon Tetrachloride	8	9	7
Chloroform	7	8	7
1,3 butadiene	6	5	4
Dichloromethane	9	9	8

Table 1. Number of Laboratories That Reported These Compounds by Year

In order to estimate the bias of the data, it was decided to compare the compounds that were reported by most of the laboratories over the three-year period and, report 1,3 butadiene, since it is a compound of interest to the NATTS.   Below are three examples from the 9 compounds that were examined (see Figures 5-7).

It can be discerned from the graphs that there are two forces at work: the standard error is decreasing and the average concentration decreased as well.   The decrease in the standard error appears to be the product of the values decreasing over the three year period and that the laboratories are better able to qualitatively and quantitatively analyze for air toxics. For instance, in Figure 7, the standard errors for Benzene in 2001, 2002 and 2003 are 0.32, 0.07 and 0.04 ppb-v respectively, which in 2003 is an 8-fold decrease in values. The means of the Benzene values for these three years are 1.53, 0.59 and 0.49 ppb-v, which illustrates that the mean value only decreased by 3-fold decrease.   This trend has been seen with most of the compounds.
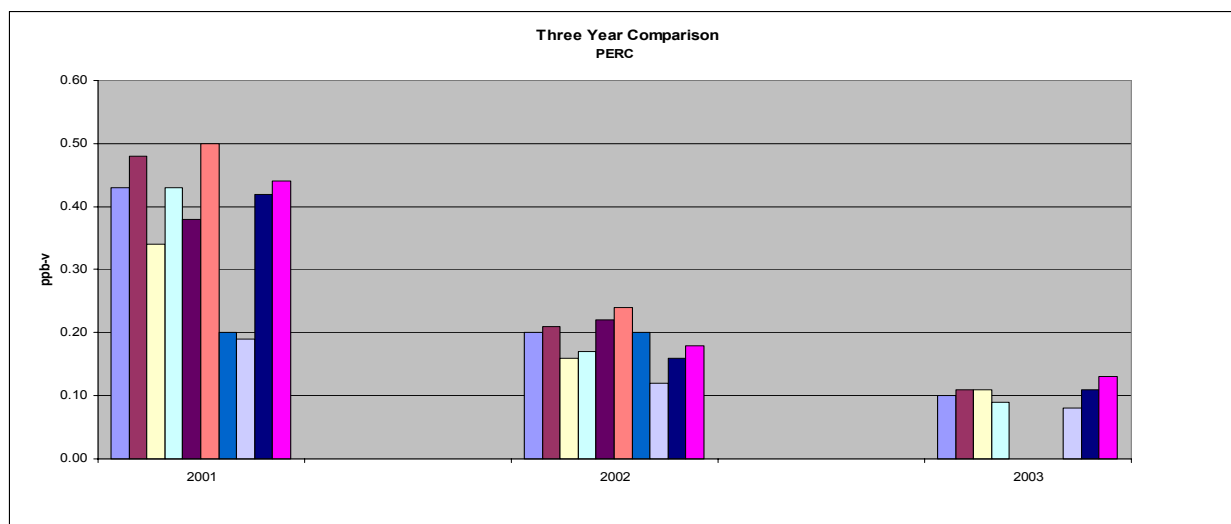

Figure 5. Three Year Comparison of Tetrachloroethylene (PERC)
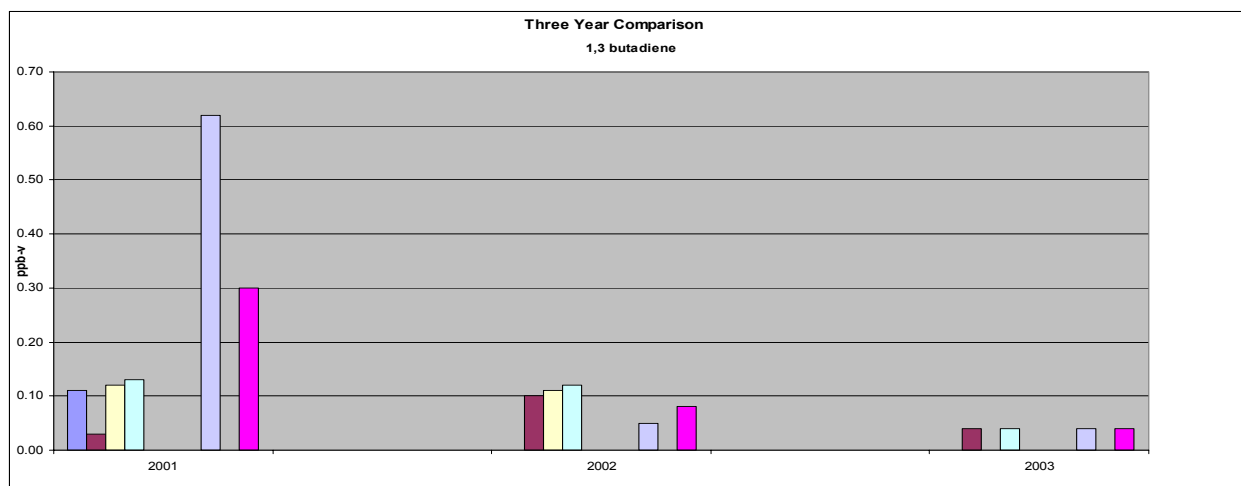
Figure 6. Three Year Comparison of 1,3 butadiene
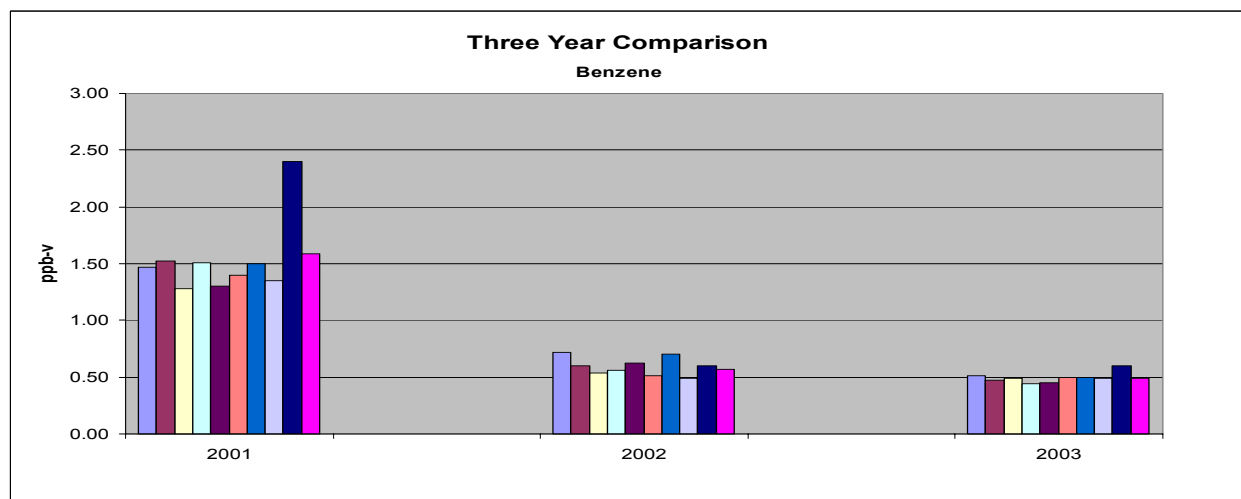
Figure 7. Three Year Comparison of Benzene

Compound	2001	2002	2003	Decrease?
Tetrachloroethylene (PERC)	0.11	0.04	0.02	Yes
Benzene	0.32	0.07	0.04	Yes
Ethyl-Benzene	0.13	0.23	0.04	Yes
m,p Xylene	0.54	1.01	0.16	Yes
Toluene	1.86	1.81	0.18	Yes
Carbon Tetrachloride	0.02	0.01	0.01	Yes
Chloroform	0.02	0.02	0.01	Yes
1,3 butadiene	0.02	0.03	0.00	Yes
Dichloromethane	0.38	0.12	0.11	Yes

Table 2. Standard Error by Year (all values expressed in ppb-v)

However, this illustrates a problem which may become apparent when the concentrations decrease as would be expected when air toxics emission controls are implemented. For example in Figure 6, for 1,3 butadiene, in 2001, there is large variability in the values. In 2002, the values even out, however, only 5 labs reported values. In 2003, only 4 labs reported data for this compound and all reported the same concentration. This raises questions concerning these values. Why did the other 6 laboratories not report any values? The author believes that the central issue may be with detection limits and an inherent problem with "Whole Air Samples." If air toxics concentrations are decreasing, as seen from these three years of data, it will be increasingly important that PT samples be created in a laboratory using National Institute of Standards and Technology (NIST) traceable gases in ranges that can easily be quantified by all laboratories in the NATTS. Therefore, as the NATTS QS moves forward, the EPA will be implementing PT samples that are NIST traceable.

References

1. Air Toxics Monitoring Concept Paper" (Draft), February 29, 2000, http://www.epa.gov/ttn/amtic/airtxfil.html

2. Quality Assurance Guidance Document -- Model Quality Assurance Project Plan for the National Air Toxics Trends Stations, 2002, http://www.epa.gov/ttn/amtic/airtxfil.html

3. National Air Toxics Trends Sites Technical Assistance Document (DRAFT), 2004, http://www.epa.gov/ttn/amtic/airtxfil.html

4. Air Toxics Pilot Laboratory Intercomparison - Quality Assurance Report, 2002, http://www.epa.gov/ttn/amtic/airtxfil.html

Multi-Site Evaluations of Candidate Methodologies for Determining Coarse Particulate Matter (PMc) Concentrations

Robert W. Vanderpool, Thomas G. Ellestad, Mary K. Harmon, Timothy D. Hanley,
Richard D. Scheffe, and Elizabeth T. Hunike
U.S. EPA, Research Triangle Park, NC 27711

Paul A. Solomon
United States Environmental Protection Agency, Las Vegas, NV 89193

Robert W. Murdoch, Sanjay Natarajan, and Christopher A. Noble
Research Triangle Institute, Research Triangle Park, NC 27709

Jeffrey L. Ambs
Rupprecht and Patashnick, East Greenbush, NY 12061

Gilmore J. Sem
TSI Inc., Shoreview, MN 55126

John Tisch
Tisch Environmental Inc., Cleves, OH 45002

Comprehensive field studies were conducted to evaluate the performance of sampling methods for measuring the coarse fraction of PM_{10} in ambient air. Five separate sampling approaches were evaluated at each of three sampling sites. As the primary basis of comparison, a discrete difference method was used which employs two designated FRM samplers, one to measure $PM_{2.5}$ and the other PM_{10}. The numerical difference of these reference method concentrations (PM_{10}-$PM_{2.5}$) represented an estimate of PMc. A second sampling approach involved a sequential dichotomous sampler, which provided both $PM_{2.5}$ and PMc measurements. In both of these filter-based, time-integrated measurement approaches, the collected aerosol mass was analyzed gravimetrically in the laboratory under controlled temperature and relative humidity conditions. Three continuous coarse particle samplers that measure PMc directly with a time resolution of 1 hour or less were also evaluated. One such sampler was a commercially available system based on beta attenuation, the second was based on TEOM technology. Both of these measurement approaches used dichotomous virtual impactors for separating fine and coarse particles. The third real-time sampler evaluated was an aerodynamic particle sizer (APS) that measures the aerodynamic diameter of individual particles, calculates the mass of the particle based on an assumed particle density, then sums the mass within the size range of interest to estimate the PMc mass concentration.
Sampling sites and timing of the studies were selected to provide diverse challenges to the samplers with respect to aerosol concentration, aerosol particle size distribution, and aerosol composition. Results from performance evaluations of the candidate PMc samplers at Gary, IN, Phoenix, AZ, and Riverside, CA are summarized.

INTRODUCTION

In response to increasing evidence of the adverse health effects associated with exposure to ambient fine particles, the United States Environmental Protection Agency (EPA) promulgated in 1997 a national ambient air quality standard (NAAQS) for $PM_{2.5}$[1]. Based on subsequent litigation, the U.S. Court of Appeals for the District of Columbia reviewed the 1997 regulations and upheld EPA's promulgation of the $PM_{2.5}$ standard but vacated the 1997 PM_{10} standard after concluding that PM_{10} is a "poorly matched indicator for coarse particulate pollution" because PM_{10} includes the $PM_{2.5}$ fraction. EPA did not appeal this ruling and now intends to promulgate a new NAAQS for PMc (i.e. the coarse fraction of PM_{10}).

Inherent to any new NAAQS is the need for sampling and analysis methods capable of measuring the new metric with known quality. In support of this goal, the purpose of this field study series was to conduct a survey of available instrumentation designed to measure the coarse fraction of PM_{10}, and to conduct a multi-site performance evaluation of these instruments. Multiple monitors of each type were used in order to determine the inherent precision of each sampler's design.

DESCRIPTION OF PMc SAMPLERS

Table 1 lists each sampler used in this study, its manufacturer, and the number of samplers used at each sampling site. For the filter-based samplers, the filter composition is listed along with the species to be determined during the filter's post-sampling gravimetric and/or chemical analysis.

Measurement Method	PM Metric	Sampler Manufacturer(s)	Number Used	Filter Composition	Species Analyzed
Integrated FRM	PM_{10}	BGI, R&P, AND	3	Teflon	Mass, sulfate, nitrate, metals
Integrated FRM	PM_{10}	BGI	1	Quartz	EC, OC
Integrated FRM	$PM_{2.5}$	BGI, R&P, AND	3	Teflon	Mass, sulfate, nitrate, metals
Integrated FRM	$PM_{2.5}$	AND	1	Quartz	EC, OC
Integrated Dichot, sequential	$PM_{2.5}$, PMc	R&P	3	Teflon	Mass, sulfate, nitrate, metals
Integrated Dichot, sequential	$PM_{2.5}$, PMc	R&P	1	Quartz	EC, OC
TEOM	PMc	R&P	3	none	none
Beta Attenuation	$PM_{2.5}$, PMc	Tisch	3	none	none

Measurement Method	PM Metric	Sampler Manufacturer(s)	Number Used	Filter Composition	Species Analyzed
Time of Flight (APS)	PMc	TSI	2	none	none
			Total = 20		

Table 1. Inventory of Samplers Used in the Performance Evaluation

Collocated $PM_{2.5}$ and PM_{10} FRM Samplers

In the first PMc measurement approach, commonly referred to as the "difference method", a designated $PM_{2.5}$ FRM sampler is collocated with a designated PM_{10} FRM sampler. For accurate determination of PMc concentrations, the PM_{10} sampler is simply a designated $PM_{2.5}$ FRM with its WINS fractionator replaced by a straight downtube. Both samplers are installed, calibrated, operated, and analyzed using standard $PM_{2.5}$ protocols. The two samplers thus have identical inlet aspiration characteristics, produce identical PM_{10} fractions, and collect aerosol at the same face velocity through the same filter media. At the completion of concurrent sampling periods, the PMc concentration is calculated as the numerical difference between the measured PM_{10} concentration and the measured $PM_{2.5}$ concentration. Due to its fundamental measurement principle, the difference method was used as the basis of comparison upon which to evaluate the performance of the other PMc samplers in the study.

R&P Model 2025 Sequential Dichotomous (Dichot) Sampler

The Model 2025 dichot was designed to provide integrated measurement of both fine and coarse fractions of a PM_{10} aerosol. The sampler actively provides volumetric flow control through a standard 16.7 actual liters per minute (alpm) PM_{10} inlet. Following the aspirated aerosol's fractionation in the inlet's internal fractionator, the resulting PM_{10} aerosol enters a virtual impactor where the aerosol is then split into major and minor flow streams. Ideally, the major flow (maintained at 15 lpm) is intended to collect only the $PM_{2.5}$ fraction of the PM_{10} aerosol while the minor flow (maintained at 1.7 alpm) is intended to collect only the PMc fraction of the PM_{10} aerosol. In practice, however, this size fractionation is never ideal and 10% of the $PM_{2.5}$ mass theoretically deposits onto the PMc filter. The presence of these fine particles is numerically accounted for during subsequent calculation of the PMc concentration. Assuming that particle losses within the instrument are negligible, the sum of the measured $PM_{2.5}$ and PMc concentrations provide a measure of the ambient aerosol's PM_{10} concentration.

Tisch Inc. Model SPM-613D Dichotomous Beta Gauge

Manufactured by Kimoto Electric Co., LTD., the Tisch SPM-613D dichot beta gauge is designed to provide near real-time measurement of both the fine and coarse fractions of the PM_{10} aerosol. The SPM-613D aspirates the ambient aerosol through a standard 16.7 lpm inlet and introduces the fractionated PM_{10} aerosol into a custom designed virtual impactor. The virtual impactor in the SPM-613D has different dimensions than that of the R&P design and operates its major and minor flow rates at slightly different flow rates, 15.2 lpm and 1.5 lpm, respectively. Downstream of the SPM-613D's virtual impactor, the separate fine and coarse flow streams are continuously collected on a paper roll composed of low hygroscopicity polyfon. Following each hour of aerosol collection, the attenuation of ^{147}Pm beta rays by each channel's aerosol deposit is quantified using two separate sets of beta sources and detectors.

R&P Continuous Coarse TEOM Monitor

The R&P continuous coarse TEOM was designed to provide a near real-time measurement of PMc concentrations. The instrument aspirates ambient aerosol through a standard 16.7 lpm size-selective inlet whose internal dimensions have been modified in an effort to provide a 10 μm cutpoint at the sampler's 50 lpm flow rate. Downstream of the inlet, the PM_{10} fraction then enters a custom virtual impactor whose major and minor flow rates are 48 lpm and 2 lpm, respectively. In this design, the fine fraction (major flow) is collected in a replaceable total filter and the collected fine aerosol mass is not subsequently quantified. Downstream of the virtual impactor, coarse aerosols in the minor flow stream are first heated to 50 °C to minimize interferences from particle bound water and are then deposited in a standard R&P 1400a Tapered Element Oscillating Microbalance (TEOM). The mass of the deposited aerosol is then estimated based on the observed change in vibrational frequency of the TEOM filter during the collection period.

TSI Inc. Model 3321 Aerodynamic Particle Sizer (APS)

The final measurement approach used in the field study involved the TSI Inc. Model 3321 APS to estimate the mass of ambient coarse particles based on their aerodynamic properties in an accelerating flow stream. In the APS, the volume of each particle is calculated based on its measured aerodynamic diameter and a particle density specified by the user. For purposes of this field study, a particle density of 2 g/cm^3 was assumed as representative for the coarse fraction of PM_{10} aerosols. The mass concentration of PMc aerosols is then calculated as the sum of the mass of all particles penetrating the PM_{10} inlet whose aerodynamic diameters were greater than 2.5 μm. Because the APS is only capable of resolving particles larger than approximately 0.7 μm aerodynamic diameter, the system is not applicable for measurement of either $PM_{2.5}$ or PM_{10} ambient concentrations because particulate mass less than 0.7 μm is not quantified.

SITE CHARACTERISTICS

The Gary, IN site was selected as representing a midwest industrial city where primary PMc aerosols are predominantly generated by industrial activity rather than by wind blown soils. Temperatures at the site ranged from -15.1 °C to 27.8 °C and a mean daily site temperature of 4.6 °C was recorded. Daily PMc concentrations (expressed as the numerical difference between

collocated PM_{10} and $PM_{2.5}$ FRM measurements), ranged from 4.5 μg/m^3 to 58.1 μg/m^3 with a measured mean of 19.8 μg/m^3. As indicated by a mean $PM_{2.5}$/PM_{10} ratio of 0.55 during the 30 sampling events, slightly more than one-half of the site's PM_{10} aerosol was associated with $PM_{2.5}$ aerosols.

Tests were conducted in Phoenix, AZ during early summer of 2003 in order to challenge the coarse particle samplers with high concentrations of dry, wind blown crustal materials. Temperatures at the Phoenix site ranged from 17.1 ^{o}C to 43.5 ^{o}C and a mean daily site temperature of 32.3 ^{o}C was recorded. PMc concentrations ranged from 26.5 μg/m^3 to 209.0 μg/m^3 with a measured mean of 55.6 μg/m^3. $PM_{2.5}$/PM_{10} ratios ranged from 0.10 to 0.28 which indicated that coarse particle mass dominated the PM_{10} concentrations during each day of the Phoenix tests.

The Riverside,CA sampling site was selected as a west coast site where significant secondary fine mode aerosols might be present in conjunction with primary coarse aerosols. Temperatures at the site ranged from 15.4 ^{o}C to 40.4 ^{o}C and a mean daily site temperature of 25.9 ^{o}C was recorded. PMc concentrations ranged from 16.2 μg/m^3 to 46.1 μg/m^3 with a measured mean of 30.4 μg/m^3. As indicated by the mean $PM_{2.5}$/PM_{10} ratio of 0.37 during the 30 sampling events, approximately two-thirds of the sites PM_{10} concentration was associated with PMc aerosols.

SUMMARY OF RESULTS

1.   Through coordination with state and local air monitoring agencies, the study sites selected met the study's siting objectives well and challenged the candidate samplers with a wide range of aerosol size distributions, aerosol concentrations, and meteorological conditions. Relatively few operational problems were experienced with the sampling equipment and the overall data capture rate for the study exceeded 95%. Prestudy, midstudy, and poststudy performance audits conducted at each sampling site revealed that the samplers typically held their calibrations well during the month-long field tests. The involvement and cooperation of the various sampler manufacturers was a key factor in the study's ability to successfully determine the inherent performance of the samplers.

2.   The filter-based, integrated samplers involved in the study provided precise test results at all three sampling sites. For the FRM samplers, the mean inter-manufacturer coefficient of variation for $PM_{2.5}$, PMc, and PM_{10} was 2.7%, 4.5%, and 2.9%, respectively. Intra-manufacturer precision of the three R&P Model 2025 dichotomous samplers for $PM_{2.5}$, PMc, and PM_{10} measurements was 2.5%, 3.0%, and 2.0%, respectively. Effective shipping protocols resulted in negligible particle loss during transport of collected aerosol samples from each sampling site to the RTP weighing facility.

3.   With the exception of Phoenix where coarse particles may have intruded into the samplers' fine channel, the R&P dichots typically provided $PM_{2.5}$ measurements which agreed closely with the collocated $PM_{2.5}$ FRM samplers. In regressions versus the collocated FRMs, all R&P dichot test results were highly correlated. The R&P dichots, however, underestimated PMc concentrations at all sampling sites with a 21% under-measurement recorded at the Phoenix site. Mass balance calculations revealed that 16%

of the aspirated PM_{10} mass in Phoenix is not accounted for during subsequent gravimetric measurement of fine and coarse channel filters. Recent tests have indicated that loss of coarse mode aerosols during the sampler's automated, post-sampling movement of the coarse particle cassette to the sample storage position may account for the observed bias.

4. In general, excellent inter-sampler precision of the R&P coarse TEOM samplers was observed at all three sampling sites and no operational problems were encountered with the samplers. However, with the exception of the Phoenix tests, the coarse TEOM tended to underestimate the PMc concentration by as much as 30%. The high correlation between the coarse TEOMs' response versus the collocated FRMs indicated that this performance was very consistent from one sampling event to another.

5. The Tisch SPM-613D samplers provided precise, highly correlated test results at all three sites for $PM_{2.5}$, PMc, and PM_{10} measurements. Although performance varied by site, the Tisch units generally provided PMc measurements within 10% of that of the collocated FRM samplers. However, the SPM-613D units consistently provided $PM_{2.5}$ concentrations significantly higher than the collocated $PM_{2.5}$ FRM samplers. As an example, the mean overestimation in $PM_{2.5}$ concentrations at the Phoenix site was 70%. Similar to the behavior of the R&P dichot, intrusion of coarse particles into the Tisch unit's fine channel may be partially responsible for this observed response.

6. With the exception of a single electronics failure, the two TSI Model 3321 units appeared to function well and provided acceptable levels of precision. Although the APS units were observed to track the PMc FRM concentrations well, they typically underestimated PMc mass concentrations by a factor of two at all sampling sites. This field behavior is consistent with previous laboratory tests of the Model 3321 conducted under controlled conditions.

DISCLAIMER

PTB Precision Results for CY2003 with Samples <6 ug/m3 Removed
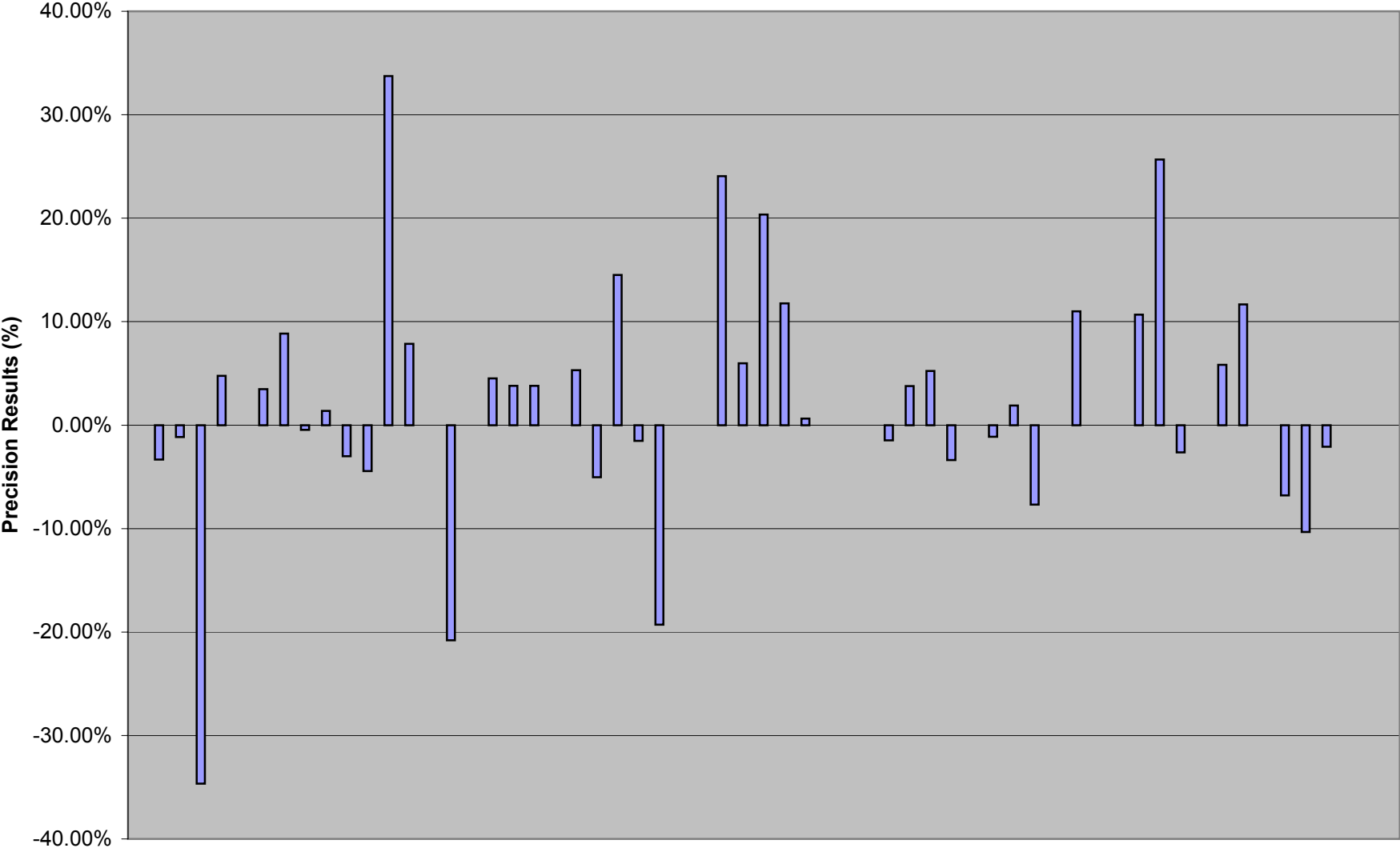
Data Quality Objective Development for the Coarse Particulate Matter Standard

Mike Papp, U.S. EPA
Basil Coutant, Batelle
S. Eberly, U.S. EPA

EPA is contemplating the development of National Ambient Air Quality Standard (NAAQS) for the coarse fraction of ambient air. This coarse fraction is generically defined as the particulate fraction from less than or equal to [a nominal]10 micrometers, to those particles greater than or equal to [a nominal] 2.5 micrometers.

Data quality objectives are qualitative and quantitative statements derived from the DQO Process that clarify study objectives, define the appropriate type of data, and specify the tolerable levels of potential decision errors that will be used as the basis for establishing the quality and quantity of data needed to support decisions.

Using some of the same techniques that were used to develop DQOs for fine particulate NAAQS (PM$_{2.5}$), the EPA developed a DQO software tool that provides decision makers with an understanding of the consequences of various input parameters, such as sampling frequency, data completeness, precision and bias and how these uncertainties affect the probability of making decision errors. Since both manual and continuos (automated) methods may be proposed for use in estimating the coarse particulate fraction, and the measurement uncertainties are unique to both methods, the DQO process can help weigh the benefits and disadvantages of these methods.

Preliminary data was collected from sites providing coarse particulate estimates from around the country as well as data from current multi-site performance evaluations conducted by the EPA National Environmental Research Laboratory. This data provided estimates of reasonable input parameters that were used to generate decision error performance curves. Preliminary decision error performance curves will be reviewed for affects of varying input parameters of precision, bias, sampling frequency and completeness on both continuous and manual methods.

This is an abstract of a proposed presentation and does not necessarily reflect the United States Environmental Protection Agency (EPA) policy. The actual presentation has not been peer reviewed by EPA. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

DQOs are qualitative and quantitative statements derived from the DQO Process that clarify the monitoring objectives, define the appropriate type of data, and specify the tolerable levels of measurement errors for the monitoring program. By applying the DQO Process to the development of a quality system, the EPA guards against committing resources to data collection efforts that do not support a defensible decision. The Office of Air Quality Planning and Standards (OAQPS) is contemplating the development of a particulate matter coarse (PM$_{10-2.5}$) National Ambient Air

Quality Standard (NAAQS). Upon promulgation of a $PM_{10-2.5}$ NAAQS, OAQPS would have a NAAQS for particles of two size fractions: those less than or equal to [a nominal] 2.5 micrometers (µm) hereafter referred to as $PM_{2.5}$; and those between [a nominal] 2.5 µm and 10 µm, defined as $PM_{10-2.5}$ also referred to as coarse particulates. Since OAQPS developed a DQO in 1997 for $PM_{2.5}$, it was felt that an effort should be made to develop a DQO for $PM_{10-2.5}$ prior to any promulgation in order to provide decision makers some idea of the potential for data uncertainty.

Decision makers need to feel confident that the data used to make environmental decisions are of adequate quality. The data used in these decisions are never error free and always contain some level of uncertainty. Because of these uncertainties, there is a possibility that decision errors can be made when measurements appear to provide an estimate above some action limit when the true estimate is below, or below an action limit when the true estimate is above. Therefore, decision makers need to understand and set limits on data uncertainties that lead to these types of decision errors. The DQO process allows one to identify these data uncertainties, determine how they affect data quality and develop quality systems and network designs to reduce or control these uncertainties to acceptable levels. The intent of this paper is to describe the process used to identify data uncertainties, and using this information, develop a DQO tool to help decision makers and those required to implement the monitoring program develop a quality system for $PM_{10-2.5}$.

The DQO Performance Curve

OAQPS used performance curves to determine the affect of various types of uncertainties on decision error. The performance curve is a powerful tool for illustrating what various uncertainties can do to the probability of making correct decisions. Figure 1 is an example of a performance curve. The terms used in the figure are explained below:

Action limit - The action limit is the concentration or value that causes a decision maker to choose one of the alternative actions. A good example of action limits are the NAAQS standards where a concentration is identified and used to determine attainment or alternatively nonattainment of the NAAQS

Performance curves - Two performance curves have been generated based upon a number of input parameters of population and measurement uncertainties. The points along the curve are the true unknown concentration. The reason for the two curves is to represent measurement bias. The curve on the left side of the action limit represents the true concentration and the decision error relative to a positive 10% bias (as well as the other uncertainty values) while the curve on the right hand side of the action limit represents a true concentration and the decision error relative to a negative 10% bias (as well as the other uncertainty values).

Decision Error Limits - These limits are established by the decision makers and presents the

decision makers "comfort" with making a decision error, in the sense that a different decision would have been made if the decision maker had access to "perfect data" or absolute truth. The decision error limit in this example is 5%.

Gray Zone - The gray zone is the area between the performance curves where the decision errors are larger than the decision error limits. This occurs because the high cost or resources required to "tighten" the gray zone outweigh the consequences of choosing the wrong course of action.

Power - This is the probability of deciding that an observed design value exceeds the action limit.
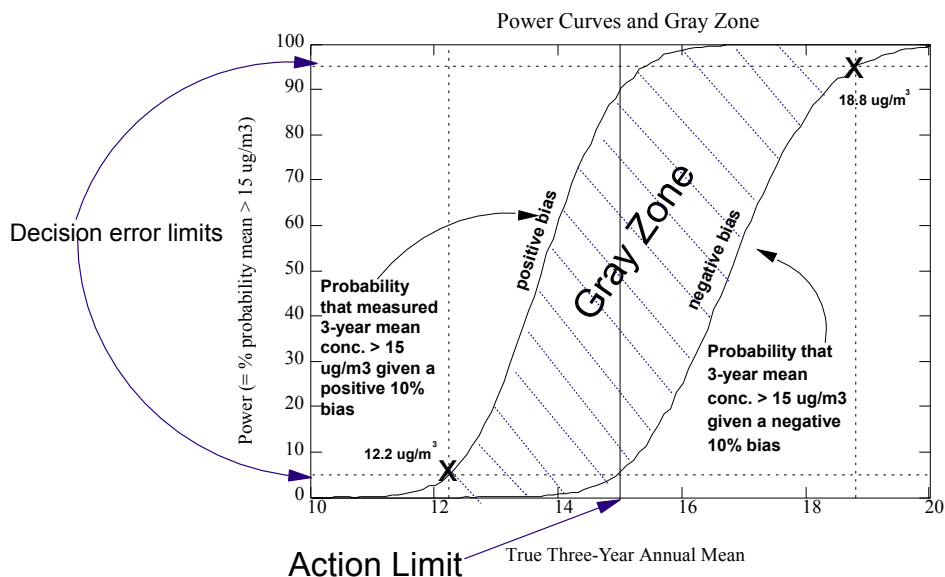
Figure 1 Example of a DQO Performance Curve

From Figure 1 the following statements could be made:

▸ If the true estimate is 18.8 ug/m^3 and if the measurement system has a negative bias of 10%, then 95% of the time the observed estimate will be above the 15 ug/m^3 action limit (correct decision) and 5% of the time the observed estimate will be less than 15 ug/m^3.

▸ If the true estimate is 12.2 ug/m^3 and the measurement system has a positive bias of 10%, then 5% of the observed estimates will be greater than 15 ug/m^3 and 95% will be less (correct decision).

160

► If bias of $\pm$ 10% is tolerable, any true estimate in the range of 12.2 to 18.8 ug/m^3 may have decision errors greater than 5%. As an example, for an estimate that truly is 17 ug/m^3 and the measurement system has a 10% negative bias, then 50% of the observed estimates will be declared to be less than the 15 ug/m^3 action limit .

The performance curve is a powerful tool for illustrating the affect uncertainties can have on the probability of making correct decisions. For example, larger biases widen the gray zone, higher data completeness narrows the gray zone. Generally, the "steeper" the performance curves or the narrower the gray zone, the higher the probability of making correct decisions around the action limit. Thus, the performance curves can identify those uncertainties that have the greatest influence on decision errors, and help focus resources to minimize those uncertainties.

Sources of Uncertainty

Decision errors can be effected by the following variables that are related to four general categories: the method, the NAAQS, the sample population or the measurement uncertainty.

Uncertainty Related to the Method

There is the possibility that both integrated manual methods and continuous methods may be used to estimate $PM_{10-2.5}$. One type of integrated method that is considered manual would require the use of two filter based sampling instruments: a PM_{10} instrument and a $PM_{2.5}$ instrument where $PM_{10-2.5}$ would be estimated by subtracting the $PM_{2.5}$ estimate from the PM_{10} estimate. Using two instruments create a potential for greater uncertainty and a widening of the gray zones. Continuous $PM_{10-2.5}$ instruments are available and have the advantage of being able to be operated every day but these instruments are still under development and display some bias in certain geographic areas. Historically, for each ambient air criteria pollutant, one method type is designated as a federal reference method (FRM). The manual methods for PM_{10} and $PM_{2.5}$ are currently designated as FRMs and may need to be used in $PM_{10-2.5}$ to provide an estimate of bias for the continuous methods.

Uncertainty Related to the NAAQS

► **Level of standard** - The level of the standard refers to the concentration where the action limit is set. For example, if an action limit is set at a concentration close to the sensitivity of the method, one would expect more potential for decision error. The information on the potential concentration ranges of the two standards is included in the Draft EPA Staff Paper: *Review of the National Ambient Air Quality Standards for Particulate Matter*. Since the standard has not been promulgated, OAQPS used the max/min of the annual and daily standard identified in the Staff Paper (see Table 1)
► **Form of the standard** - If one uses an annual average versus the highest concentration in a year, there would be more potential for decision error with the single high concentration value.

Current thinking on the $PM_{10-2.5}$ is to propose two standards similar to the current $PM_{2.5}$ standard; a three year annual average value (annual average) and a 3-year percentile of a 24 hour average value (daily standard). OAQPS developed DQO scenarios for both forms.

▶ **Percentile for daily standard** - different percentiles of the daily standard could affect decision error. OAQPS looked at 98, 95 and 90 percentiles of a 3-year 24 hour average but did not notice significant differences in the DQO gray zone and so used a 98 percentile.

Uncertainty Related to Sample Population

Values related to sample population were developed through a data review of PM_{10} and $PM_{2.5}$ data available in AQS. Values for each attribute were selected at a conservative but realistic level, meaning 90-95% of the sites had values less than the ones chosen for input to the DQO performance curves. Population uncertainty inputs are not changed when running DQO performance curves scenarios.

▶ **Seasonality ratio** - is the ratio of the highest concentration to the lowest concentration within a particular time period. A ratio of 7 for $PM_{10-2.5}$ was used.
▶ **Population variability** - this measures the random, day-to-day movement of the true concentration about the average sine curve. 60% for $PM_{10-2.5}$ was used.
▶ **Autocorrelation** - is a measurement of the estimate similarity on successive days. Since there is a possibility that $PM_{10-2.5}$ can be measured on a 1 in 6 day sampling frequency, an autocorrelation of 0 was used. If continuous instruments are used, everyday sampling will be viable and some autocorrelation may be incorporated into the DQO.

Uncertainty Related to Measurement System

▶ **Sampling frequency** - The DQO tool used both 1 in 6 day and every day sampling frequency to accommodate both manual and continuous methods.
▶ **Completeness** - 75% was used since it is currently allowed in CFR for particulate matter.
▶ **Measurement bias** - 10% bias was used since this appears reasonable for $PM_{2.5}$ and would probably remain reasonable. More information on this uncertainty is being assessed.
▶ **Measurement precision** - 10% precision was used since this appears reasonable for $PM_{2.5}$ and would probably remain reasonable for either a manual or continuous method. More information on this uncertainty is being assessed.

The DQO Software Tools

The DQO tools use performance curves which allows one to model $PM_{10-2.5}$ data based on the fixed population uncertainty assumptions. Then, the performance curves are changed based on the inclusion of measurement uncertainty input parameters of sampling frequency, precision, bias and completeness. The goal is to keep the gray zone as narrow and the performance curves as steep as possible. Two DQO software tools were developed: one, the direct measurement tool, that can be
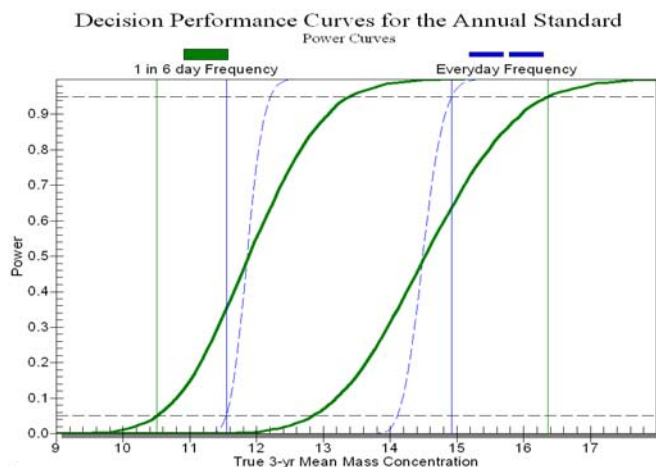
Figure 2 Direct DQO Tool Example

used for continuous instruments or manual instruments that provide $PM_{10-2.5}$ material on a single filter; and a second tool, the integrated tool, when the method requiring a PM_{10} instrument and a $PM_{2.5}$ instrument, is used. The DQO tools allow one to generate gray zones for a number of scenarios in which different measurement uncertainty values are altered and these curves can be displayed on the same graph. By altering these uncertainty values, one can determine which uncertainty has the most affect on data quality. Figure 2 provides an example graph derived from the direct DQO tool where only sampling frequency was altered from 1 in 6 day to everyday.

Performance Curve Results

Table 1 provides gray zones for $PM_{10-2.5}$ at the NAAQS levels mentioned in the Draft Staff Paper. This table provides an example of the changes in the width of the gray zone in relation to sampling frequency and sampling method. For $PM_{10-2.5}$ the gray zone estimates for columns 3 (1-6 day integrated) and 4 (every day integrated) were developed from the DQO software where two instruments (a PM_{10} and a $PM_{2.5}$) are used to derive a $PM_{10-2.5}$ concentration. The 5th column, identified as "direct" is the gray zone derived either from a continuous instrument or an instrument collecting a coarse sample on one filter. The reason for the larger gray zones from the integrated method even when every day sampling occurs are related to the additive errors of two methods in order to derive a concentration.

Table 1. PM $_{10-2.5}$ Gray Zone Performance Curve Values Relative to Sampling Frequency and Method

Standard	NAAQS ug/m³	1-6 Day (Integrated)	Every Day (Integrated)	Every Day (Direct)
Annual	13	7.8 - 26.1	8.3 - 23.4	11.5 - 15.0
	30	18 - 60	19.2 - 53	26.6 - 34.4
Daily 98%	30	14.5 - 46.4	19.2 - 42.7	25 - 36.4
	75	35 - 113	47.7 - 105.8	62.1 - 89.8

Conclusions

The DQO software provides user-friendly insights into the affects of uncertainty on decision making and identified that the annual standard gray zones are most sensitive to population variability, sampling frequency, measurement bias, and completeness. The daily standard is sensitive to the variables listed above in addition to precision. Results from the DQO work are preliminary. The information on the form and the level of the standard are draft proposals and not certain and are used only to provide an example of the DQO software's capability. The population and measurement uncertainty parameters have not been agreed upon and may change, thus changing the gray zones. The EPA National Environmental Research Laboratory is currently conducting intercomparsons on a number of the $PM_{10-2.5}$ manual and continuous instruments. This information will be used to check the population and measurement uncertainty assumptions in order to revise the software as needed and to provide more accurate assessments of the potential for decision errors**.**

New Quality Indicator Statistics for the Gaseous Criteria Pollutants

Basil Coutant, Battelle
505 King Avenue
Columbus, OH 43201
coutantb@battelle.org

State and local agencies are responsible for the routine sampling of gaseous criteria pollutants, and the U.S. EPA coordinates these programs. A key part of this coordination is providing a framework for data quality objectives (DQOs) and the evaluation of data quality indicators. The framework for the proposed ozone DQOs is similar to $PM_{2.5}$ in that it is based on a minimum sampling frequency, a minimum completeness rate, a maximum allowable absolute bias (a measurement of consistent measurement error), and a maximum allowable precision (a measurement of random measurement error). DQOs for all the gaseous criteria pollutants will likely be based on the ozone DQO framework.

The sampling frequency and completeness requirements and quality indicator statistics are unchanged from previous requirements and statistics. The absolute bias and precision are new and have new statistics associated with them. The new quality indicator statistics are based on quality indicator data already reported in AQS; they do not require the collection of additional data. They are designed to make better use of the data already collected.

Both the precision and bias are based on the relative error between the measured concentration and the "known" concentration of a test gas whose concentration is within a range specified in the Code of Federal Regulations (CFR). Checks against a known gas are required at least biweekly and are commonly known as biweekly precision checks.

The proposed precision indicator is a statistical upper bound on the standard deviation of the relative errors. The absolute bias indicator is an upper bound on the mean absolute values of the relative errors. The old probability intervals are still computed as before with these data. They are now compared to the "accuracy audit" relative errors by verifying that at least 95 percent of the accuracy audit relative errors are within the probability intervals. Together these statistics provide a comprehensive evaluation of the data quality. The paper describes both the details of the new statistics and some of the rationale behind the statistics.

Introduction

State and local agencies are responsible for implementing programs to routinely sample criteria pollutants [5]. The U.S. EPA[5] coordinates these programs by providing a framework for both the data quality objectives (DQOs) and the evaluation of data quality indicators. The EPA also

maintains the AQS database, which stores a permanent record of the data and generates summary reports of the data. EPA has noted that the current statistics being calculated from data quality indicators do not directly apply to the DQOs. Since the DQOs are typically developed using a bias parameter to measure the magnitude of the error and a coefficient of variation (CV) parameter to measure the random component of the error, a need exists for more relevant, robust statistics for these two components.

Table 1 summarizes the main quality indicator data collected in AQS for each of the criteria pollutants and the various summary statistics.

Table 1. Current QA Summary Statistics

Pollutant	Data	Current Summary Statistic(s)
NO_2, SO_2, CO, and O_3 (Automatic methods.)	Biweekly Precision Checks Annual Accuracy Audits Span Checks[1]	Probability Interval Probability Interval None
Lead	Flow rate audits Lead strip audits Co-located measurements	Probability Interval Probability Interval Probability Interval
PM_{10}	Flow rate audits Co-located measurements	Probability Interval Probability Interval
$PM_{2.5}$	Flow rate audits Co-located measurements PEP measurements[2]	Mean percent deviation from target flow. CV estimated using the Root-Mean-Square of percent differences. Bias estimated by mean of percent differences.

[1]   Span check data are not required to be submitted to AQS, but some agencies do submit these data.
[2]   Only the state data are currently housed in AQS.

For the gaseous pollutants the main quality assurance tools are the "biweekly" precision checks. The precision checks are made by sampling from air with a known concentration of a given pollutant. A probability interval based on the relative percent error of these checks is created. This probability interval is the main method for summarizing the relative percent errors and serves well as a summary tool. (See the following section for details on calculating relative percent errors and probability intervals.) It provides a robust method based on all the data available that should, in most cases, indicate the range for 95 percent of the relative errors (both measured and unmeasured). This interval is used to control data quality by controlling the extremes in the relative errors. However, the DQO process has shown that the relative importance of random errors (precision) and persistent/consistent measurement errors (bias) is not the same for decisions based on aggregates of the data. Consequently the DQO statements for ozone, $PM_{2.5}$, and soon the other pollutants control these two components separately. Hence the single probability interval does not relate directly to the DQOs. This paper presents a set of statistics that relate directly to the DQOs and are motivated, in part, from the probability interval. They are based on the data already collected and, moreover, can be consistently extended across all the criteria pollutants.

Recommendations

Many of the following recommendations have been presented to an EPA workgroup [2] along with examples to demonstrate the strengths and weaknesses of these and other statistics. To the extent possible, the recommendations are either those of the workgroup or variants based on the approach taken by the workgroup.

The key recommendation for gaseous pollutants is that precision and bias should be estimated from the biweekly precision checks described above (henceforth referred to as "single-point" checks). As required by the Code of Federal Regulations (CFR) [5] the concentration of the reference pollutant must fall within a narrow range so both bias and repeatability precision can be estimated (see Table 2). The recommended bias statistic is an upper bound on the mean absolute value of the relative percent errors (Equation 2), and the recommended precision statistic is an upper bound on the standard deviation of the relative percent errors (Equation 4).

For each single-point check, calculate the relative percent error, d,

$$d = \frac{ind - act}{act} \cdot 100 \qquad \text{(Eq. 1)}$$

where ind is the concentration indicated by the agency's measurement and act is the actual concentration being measured. Again the data would be considered valid only if the actual value is in the range specified in CFR (see Table 2).

Table 2. CFR Single-Point Audit Range Restrictions.

	CO	NO$_2$	O$_3$	SO$_2$
Single-point check range (PPM)	8-10	0.08-0.10	0.08-0.10	0.08-0.10

The bias of the site for an aggregation period is controlled by controlling an absolute bias upper bound. The statistic (for any level of aggregation) is calculated as follows:

$$\left|bias\right| = AB + t_{0.95,n-1} \cdot \frac{AS}{\sqrt{n}} \qquad \text{(Eq. 2)}$$

where n is the number of single-point checks being aggregated; $t_{0.95,n-1}$ is the 95th quantile of a t-distribution with n-1 degrees of freedom; the quantity AB is the mean of the absolute values of the d's and is calculated via:

$$AB = \frac{1}{n} \cdot \sum_{i=1}^{n} \left|d_i\right| \qquad \text{(Eq. 3)}$$

and the quantity AS is the standard deviation of the absolute value of the d's and is calculated by:

$$AS = \sqrt{\frac{n \cdot \sum_{i=1}^{n} \left|d_i\right|^2 - \left(\sum_{i=1}^{n} \left|d_i\right|\right)^2}{n(n-1)}}. \qquad \text{(Eq. 4)}$$

The precision of the site for an aggregation period is controlled by placing bounds on a coefficient of variation upper bound, which is calculated using the formula:

$$CV = \sqrt{\frac{n \cdot \sum_{i=1}^{n} d_i^{\,2} - \left(\sum_{i=1}^{n} d_i\right)^2}{n(n-1)}} \cdot \sqrt{\frac{n-1}{\chi_{0.05,n-1}}} \qquad \text{(Eq. 5)}$$

where $\chi_{0.05,n-1}$ is the 5th percentile of a chi-squared distribution with n-1 degrees of freedom.

The absolute bias upper bound and CV upper bound should be calculated quarterly at the site level. These quarterly values, along with annual and two- or three-year values, should be compared with DQO targets for the absolute bias and precision.

The current accuracy audits can be used to verify the results obtained from the single-point checks and to validate those results across a range of concentration levels. Annual and three-year agency-level probability limits calculated from all the single-point checks should capture approximately 95 percent of the relative percent differences from the accuracy audits (for all levels) and are calculated by:

$$\textit{Upper probability Limit} = m + 1.96 \cdot S \qquad \text{(Eq. 6)}$$
$$\textit{Lower probability Limit} = m - 1.96 \cdot S \qquad \text{(Eq. 7)}$$

where *m* is the mean:

$$m = \frac{1}{k} \sum_{i=1}^{k} d_i \qquad \text{(Eq. 8)}$$

k is the total number of single-point checks for the year and agency, and *S* is the standard deviation of the relative percent differences:

$$S = \sqrt{\frac{k \cdot \sum_{i=1}^{k} d_i^{\,2} - \left(\sum_{i=1}^{k} d_i\right)^2}{k(k-1)}} \cdot \qquad \text{(Eq. 9)}$$

If the relative percent errors from any of the accuracy audits fail to be within this interval, additional QA checks should be made.

Finally, it may be desirable for a sign to be associated with the absolute bias for a site (indicating positive bias or negative bias). The recommendation for how to make this association depends on the goals for assigning a sign to the bias, which are not entirely clear from workgroup discussions. First, if the goal is to flag all sites with whichever tendency is "stronger," then the absolute bias upper bound should be flagged with the sign of the mean, sign(*m*), in Equation 8. If the goal is to flag potential problems, then sites should be flagged only when a clear majority

of the relative percent errors have the same sign. A simple algorithm for this would be to calculate the 25th and 75th percentiles (the first and third quartiles) of the relative percent errors for each site. The absolute bias upper bound should then be flagged as positive if both of these values are positive. The absolute bias upper bound should be flagged as negative if both of these values are negative. The absolute bias upper bound would not be flagged if the 25th and 75th percentiles are of different signs.

Discussion

The recommended statistics for bias and precision are not estimates of the precision and bias, but rather are statistics used to control the precision and bias. The DQOs for ozone and $PM_{2.5}$ and the DQOs being developed for the other criteria pollutants are made under the assumption that the absolute value of the bias is never more than the DQO value and that the precision is never more than the DQO value [3]. Consequently, the statistics used for quality control need to control the range of the bias and the precision. In this sense the new statistics are much like the probability interval that was used to control the range of the relative percent errors.

Bias
Since the bias is the more influential of the two types of error on decision quality, the bias is the more strongly controlled under the scheme. The statistic in Equation 2 has two conservative components. The first is that it is based on absolute values and the second is that it uses a confidence limit upper bound. The absolute values were chosen to detect or control for cases where the bias is positive part of the time and negative part of the time. Seasonal biases can cause this problem. No examples of seasonal biases were found with ozone or FRM $PM_{2.5}$ samplers (although some continuous $PM_{2.5}$ samplers show this problem). This problem also appears when bias estimates are aggregated across sites. Examples of this were found where one or two ozone sites have clearly different biases from the rest of the reporting agency. The absolute bias upper bound is better at detecting this behavior, while showing that the vast majority of the ozone sites do meet the DQO requirements [1].

The use of a confidence limit upper bound adds an additional protection, in this case, against random errors in the estimate of the mean of the absolute relative errors. It is not as conservative as might be expected at first glance. First, the upper bound is from a confidence limit, rather than a probability interval (note the square root of n in Equation 2). This is used because we are trying to control the consistent portion of the error only rather than the extremes in the absolute relative errors. Second, the standard deviation used in Equation 4 is the standard deviation of the absolute values. This is less than the standard deviation of the signed values in Equation 9.

Precision
Equation 5 is a standard confidence limit upper bound for a standard deviation [4]. The confidence limit upper bound protects against random errors in the estimate of the standard deviation. The DQO quantity of interest is the CV of the measurement error, so it would not be appropriate to use the standard deviation of the absolute values as in the bias statistic. The statistic is less conservative than the root-mean-square statistic currently used for $PM_{2.5}$, because it includes a mean correction (the second term under the first square root). This was felt to be

appropriate for the gaseous pollutants. Moreover, the statistic in Equation 5 is being considered as a replacement for the current statistic used for precision in the $PM_{2.5}$ program. So, the two precision statistics should be consistent in the future.

Confidence Intervals
The final recommendation is to use the "accuracy audits" as a means of checking assumptions made with the single-point checks. These accuracy audit checks are made too infrequently to be the basis of estimating precision or bias at a site or reporting agency. However, they do provide an excellent means of verifying a key assumption made with the single-point checks. The single-point checks are useful for estimating precision and bias because they are done with a standard that has a tightly controlled range of concentrations near the action limits. (See Table 2.) However, to be useful, the error structure needs to hold throughout a range of concentrations that includes not only the range of measured concentrations at the site but also the action limit. The accuracy audits provide this check. The probability interval in Equations 6 and 7 generated from the single-point check values should contain the relative errors found during an accuracy audit. This is not a complete check of the assumptions, but it is a useful one that can be done with the data already collected.

Conclusions

The statistics presented make better use of the QA data currently collected by the State and Local agencies monitoring the gaseous criteria pollutants. They separately control the precision and bias as required by the DQO statements. They are not estimates of precision and bias, but rather upper bounds to control the bias and precision.

References

[1]   United States Environmental Protection Agency, "National Ambient Air Quality Standards for Particulate Matter; Final Rule," 40 CFR Part 58, *Federal Register*, July 1, 1999.

[2]   Coutant, B. and Holloman, C. "Statistics for Criteria Pollutant Data Quality Objectives," draft technical report to the U.S. Environmental Protection Agency Office of Air Quality Planning and Standards under Contract No. 68-D-02-061, September 2003.

[3]   Coutant, B. "Ozone Data Quality Objective Development," *Proceedings EPA 22nd Annual National Conference on Managing Environmental Quality Systems*, New Orleans, Louisiana, April 2003.

[4]   Coutant, B., Chang, O., Biddle, B., Paul, D., Slone, E., and Young, G. "2000-2002 Criteria Pollutant Quality Indicator Summary Tables and Figures, NAMS, SLAMS, and Tribal Sites," draft technical report to the U.S. Environmental Protection Agency Office of Air Quality Planning and Standards under Contract No. 68-D-02-061, January 2004.

[5]   Freund, J., and Walpole, R. <u>Mathematical Statistics, Third Ed</u>., Prentice-Hall,
        Engelwood Cliffs, N.J., 1980.

Status and Changes in EPA Infrastructure for Bias Traceability to NIST

Mark Shanis, Environmental Scientist, U.S. EPA OAQPS

Changes are occurring in a number of the parts of the EPA infrastructure authorized and established in the 1980's to characterize and promote traceability of EPA ambient air monitoring data to NIST standards, the basis of the centralized, comparable accuracy of data in the USA. These benchmark Quality Assurance (QA) programs support staff training and evaluation of the comparability of the calibrations that all reporting organizations use to assign values to the otherwise undefined instrumental signals that air monitors provide and are the initial basis of the data reported to EPA for compliance and other purposes.

This discussion will address status and changes in the EPA's National Performance Audit Program (NPAP) for Ambient Air Criteria and other Pollutants, the EPA's Standard Reference Photometer (SRP) Program for traceably standardizing ambient ozone measurements, and the EPA Protocol Gas verification by independent, EPA-approved, third Party.

In 1996, divested itself of EPA's Office of Research and Development of its QA service programs and our office agreed to assume the responsiblities, depending on resources allocated and mission priorities.

The NPAP program changed in 1999 with the addition of portable, collocated, PM2.5 samplers, which were delivered operated, and retrieved by a nationally coordinated but Regionally based and supervised contractor. This effort was funded with State and Local Grants (STAG). An effort was started in 2001 to improve the mailed, non-PM 2.5 NPAP by combining it with the PM2.5 Performance Evaluation Program (PEP), as NPEP. The first step was to provide a network of mobile audit laboratories, based in each EPA Region, operated by EPA or PEP contractor staff, which could deliver laboratory quality audit gases and the collocated samplers. The mobile laboratories would have the capacity to verify the gases on site, and then deliver them to the inlet of the audited station, instead of just to the back of the audited analyzer. Trailer vs. truck and implementation trade-offs and approximate costs are are discussed.

The mobile laboratories are designed to allow transport and deployment of the PM 2.5 devices, as well as of other audit equipment, and of emergency hazardous (toxic or other) air sampling, of short duration, when Regional priorities dictate. This flexibility of design also addresses the need to acknowledge that monitoring technology develops and evolves, and therefore so does the need for audit equipment, methods, and infrastructure support. As of the meeting, 6 equipped mobile audit laboratories have been delivered to 6 EPA Regions (2, 4, 5, 6, 7, and 9). Three performed TTP evaluations in 2003. All 6 are expected to deliver performance evaluation samples in CY 2004.

The SRP network of 10 NIST manufactured and certified systems are deployed, based, and operated in 8 of the 10 EPA Regions. Initially the regional systems were certified as traceable to NIST using two SRPs, based in RTP, and now in Las Vegas. The other SRPs

are operated by EPA Regional staff, with one exception. The EPA Region 9 SRP has been operated by staff of the California Air Resources Board (CARB). In the last 3 years, the network has undergone one upgrade and is in the process of a second. Those SRPs that have successfully had both upgrades are able to automatically perform the documented SRP procedure and record the generated data used for certifying itself against the coordinating SRP, and to certify primary and/or transfer standards from state and local agencies, and approved EPA contractors.

Due to problems with the reliability/variability of the vendor-certified accuracy of the standard gases used to calibrate ambient air gaseous monitors, EPA established, modified, and expanded the scope of its Traceability Protocol for Certifying Gaseous Calibration Standards. In the early 1990s, EPA ORD started reporting the results of a relatively small Protocol Gas verification program. Although the sample size was small, vendors paid attention-for very low cost, results improved over the 4-5 years of the program. After ORD's divestment, the program was not continued, but EPRI (ca.1998), and EPA (2003), in response to complaints by individuals from the user community, each performed an additional blind sampling study and found that without the program, significant problems, across pollutants, had again occurred. EPA is therefore looking into a vendor-funded, EPA-approved, 3rd party-operated, blind sampling, publicly-reported verification program.

This discussion will address status and changes in the EPA's National Performance Audit Program (NPAP) for Ambient Air Criteria and other Pollutants, the EPA's Standard Reference Photometer (SRP) Program for traceably standardizing ambient ozone measurements, and the EPA Protocol Gas verification by an independent, EPA-approved, third Party.

Changes have been and are occurring in a number of the parts of the EPA infrastructure authorized and established in the 1980's to characterize and promote traceability of EPA ambient air monitoring data to NIST standards, the basis of the centralized, comparable accuracy of data in the USA. These benchmark Quality Assurance (QA) programs support staff training and evaluation of the comparability of the calibrations that all reporting organizations use to assign values to the otherwise undefined instrumental signals that air monitors provide and are the initial basis of the data reported to EPA for compliance and other purposes.

Importance of the Functions of the Traceability Infrastructure
The Traceablity infrastructure has two important functions. The first role is to give EPA Regional oversight managers a handle on S&L proficiency, especially when used in combination with TSAs. Second, probably the greatest value of the NPAP, SRP, and Gas Protocol Verfication is to provide state and local agency managers with an independent benchmark tool to check the cost effectiveness of their ongoing operator training, procedure review, data validation, equipment maintenance, and calibration standard recertification activities in their organization.

These training and benchmark roles of the comparability infrastructure verfication tools (including systems audits) are especially important for two critical reasons:

1) The turnover of government environmental positions such as field operators and lab analysts in ambient air monitoring is normally a problem, but it has been escalating around the country as the Clean Air Act anniverary approaches its 35th year, as well as because of national priorities.

2) The national level benchmark is important because 40 CFR Part 58, does not provide any other independent mechanism for determining how well agencies are doing in carrying out the requirements for the quarterly reporting of annually required, agency-funded audits.

Background/Status of NPAP
Since 1979, participation in the NPAP has been a QA requirement (40CFR part 58, appendices A, B, and C). Devices or materials have been provided as single blind samples used to evaluate the proficiency of the performance of EPA-required methods by the state local, or private monitoring station operators (and their equipment, standards, procedures, management,etc.). Some of the audits are of lab proficiency only, and some test field sampling and lab analyses and reporting. All audits are performed by the audited agency staff, usually the station operator. All audits are provided by a single, EPA-supervised audit support contractor. As monitoring equipment used in the field have evolved from wet chemistry to continuous methods, so have the audit methods. Unfortunately, recognition of and provision for this evolution has not been built into the regulations or the supporting funding mechanisms.

RTP QA Changes-ORD to OAQPS
EPA's Ambient Air QA program started changing in 1996 when EPA ORD divested itself of its QA service programs and EPA OAQPS agreed to take over, as well as it could, depending of course every year on resources allocated and mission priorities.

NPAP Changes-Creation of PEP
The particulate portion of the NPAP started changing with the addition in 1999 of portable, collocated, PM2.5 samplers, delivered, operated, retrieved and reported by a nationally coordinated, regionally based, EPA supervised, contractor. This program was first approved following many months of communications and eventual agreement between U.S. EPA OAQPS and almost all of the state and local ambient air monitoring agencies. The program is funded with State and Local agency Grants (STAG, 103 type), and is called the Performance Evaluation Program (PEP). Documents and reports of this program are available through the website for ambient monitoring.The website's URL is: http:\\www.epa.gov\ttn\amtic

An effort was started in 2001 to improve the non-PM2.5 NPAP by combining it with the PEP, as NPEP, by adding a system of Regional mobile audit laboratories. These laboratories would be to based in the EPA Regions, as is the PEP program. The mobile laboratories would provide laboratory quality audit gases verified at the audit site, and then delivered through the sampling inlet, or probe, and multi-instrument sampling manifold of the audited station.

Most audits in the US, including the mailed NPAP audits and the agencies' own quarterly reported audits, are delivered just to the back of the audited analyzer, bypassing station inlet, manifold, and connecting tubing. The model is the California Air Resources Board's (CARB's) Through-the-Probe (TTP) Mobile Audit Program. It has been in operation for about 20 years, is documented on the CARB website. The SOP for the CARB TTP program is included as Appendix in the EPA QA Handbook, Vol II, Part 1, which is posted on the AMTIC website.

There are a number of important technical and quality differences between the existing mailed EPA NPAP and the CARB audit program's capability and features. The CARB and the new EPA

TTP systems are transported, delivered under very favorable environmental conditions, and operated and reported by trained 3rd party staff, completely independent of the agency being audited. The mailed NPAP equipment has to be shipped. Therefore it is as compact and low in weight as possible, for lower shipping costs, more rugged, and therefore potentially less sensitive (precise and/or accurate) as hand-carried equipment.

The mailed equipment has the following constraints:

1) limited audit gas volumes to a few liters per minute, just enough for audits at the back of one analyzer - at the most 2- at a time. The 2 CARB systems and the 6 EPA TTP systems generate 15 lpm, or more, if desired. The current TTP zero air generator's supply limit is 20 LPM.The current calibrator's upper capacity is 40 LPM. Since a 2nd zero air system will be available shortly for the network operators to use as needed, soon the uppoer limit will be 40 LPM. Most gas genreatjkon sytem upper cpascites are based on the sampling flow rate requirements at the back of individual sampler/analyzers, which is 1-2LPM. The immediate consequence, and purpose, of this almost ten-fold higher flow rate (from 1or2 up to 20LPM) is to generate enough capacity to provide an excess of flow, at just about ambient pressure, to meet the typical CA ARB and national station inlet sampling rate of not more than approx.12-14 lPM. In that way, all surfaces along the whole flow path can be tested by the audit, and not just those from the back of the analyzer (BOA).

2) The best, most sensitive technology allows for onsite, pre-audit verification, to within a few % of the NIST-traceable standards on board the lab, as well as trouble-shooting capabilities in case problems are found and resolution priorities allow the additional time to be spent. Other important features of the system: 150 ft delivery hose, and adequate zero air generation pressure capacity to push the air along that length without the development of impeding back pressure/flow resistance. Line loss for ozone is typically no more than 2%.

Based on the currently smaller number of pollutants audited at the national level, and on the experience and advice of the CARB Audit staff leaders, the EPA TTP mobile audit lab design was reduced from CA ARB's two, 30 ft long exterior, van-based lab bodies to 5 trailer bodies of 18 foot length, and one enclosed cab and chassis cutaway truck of a 16 foot body length (plus another approx 6-7 feet for the cab, and tow vehicles for the trailers). Region 4, perhaps one of the largest and population-dense Regions, and containing the largest number of states, required a self-propelled lab Tha is also what CARB uses, for greater ease of maneuvering and driving. Five other EPA Regions opted for the greater flexibility and significantly lower cost of a trailer-based lab, with tow vehicle when needed.

The mobile laboratories are designed to allow transport and deployment of the PM 2.5 devices, as well as other audit equipment, and of emergency/hazardous air sampling, of short duration, when Regional priorities dictate. For example, the roof sampling platform feature, with collapsible guardrails and associated roof duplex receptacles, is currently on three of the six mobile labs. The flexibility is enhanced by the expanded or expandible capacity of the data logging system, digital connectivity, instrument rack, UPS-PLC system and power source system features. This flexibility of design also addresses the need to acknowledge that monitoring technology develops and evolves, and therefore so does the need for audit equipment, methods,

and infrastructure support. As of this meeting, 6 equipped mobile audit laboratories have been delivered to 6 EPA Regions (2, 4, 5, 6, 7, and 9). Three performed TTP evaluations in 2003. All 6 are expected to deliver performance evaluation samples in CY 2004.

Summaries of regionally-based TTP audit results and program cost data collected from CY 2004 activities will be communicated to the EPA Regions and to the state and local agencies. These agencies will be asked, based on this information, to agree to request grant funds to pay for the EPA PEP and the NPEP TTP audits. This approach was based on the feedback obtained in conference calls in 2001 with EPA Regional representatives, regarding improvements needed and desired in the NPAP, and at a STAPPA/ALAPCO meeting in Chicago, in May, 2001. Attendees agreed that, in concept, combination of the mailed NPAPand independently delivered PEP programs is a more cost-effective use of the NPAP and PEP-ESAT funds than is now allowed with the separately operated and funded programs. Before agreeing to request the combination, attendees wanted to see the results of pilot of the program.

Net NPAP Status and Changes
The cost of providing the equipment for the first four new regional systems initially came from a one-time OAQPS NPAP improvement initiative fund of $375,000. The rest of the funds, needed for both the remaining equipment and the contract labor to operate and maintain the current total of 6 systems, has come from the contract S&T funds that have been provided by OAQPS in 2002-2004. The number of NPAP mailed audits that the 10 Regions had become used to receiving, without more than the agency cost in performing the audits, has dropped drastically. This reduction had already started before the Regional TTP system development started, as a result of a reduction in NPAP contract funds that started in 1999, due to competing program priority needs, and has continued since then. As the 6 Regional PEP+TTP-NPEP systems, and the remaining complementary mailed NPAP program, show what they can do, for the funding amounts they have been given, the number and quality of audits will increase to a more effective level during the remainder of 2004.

EPA SRP Network Status and Changes
The SRP network of 10 NIST manufactured and certified systems are deployed, based, and operated in 8 of the 10 EPA Regions. They were coordinated (certified as traceable to NIST) initially by the relatively standard combination of a stationary and a traveling SRP, both based in RTP. Currently the primary (coordinating, traveling) SRP is based in Las Vegas. The network of Regional SRPs has, all along, and is currently operated by EPA Regional staff, with one exception. The EPA Region 9 SRP has been operated by CARB.
In the last 3 years, the network has undergone one upgrade, and is in the process of a second. Each change had improved both hardware and software. Those SRPs that have successfully had both upgrades are able to do what they could not do before, which is to automatically perform the documented SRP procedure and record the generated data used for certifying itself against the coordinating SRP, and to certify primary and/or transfer standards from state and local agencies, and approved EPA contractors. These upgrades have brought the systems up to date with the improvements of newer hardware and software that have been included in the systems assembled since the last EPA SRP was made in1989.

NIST STANDARD REFERENCE PHOTOMETERS

SRP#	COMPLETION DATE	LOCATION	ORGANIZATION
0	Aug.27, 1985	Gaithersburg, MD	NIST ("Backup")
1	Feb. 9, 1983	Raleigh, NC	EPA (ORD Lab; "Backup")
2	Feb. 9, 1983	Gaithersburg, MD	NIST ("Primary")
3	Aug. 23, 1983	Edison, NJ	EPA Region 2
4	Sep. 16, 1983	Sacramento, CA	EPA Reg. 9 (CA ARB)
5	March 20, 1985	Houston, TX	EPA Region 6 (Lab)
6	March 7, 1985	Chicago, IL	EPA Region 5
7	Jan. 23, 1986	Las Vegas, NV	EPA ORIA ("Primary")
8	Feb. 11, 1986	Denver, CO	EPA Region 8
9	May 6,1987	Lexington, MA	EPA Region 1 (Lab)
10	Nov. 4, 1987	Athens, GA	EPA Region 4 (Lab)
11	Sep. 25, 1987	Nyköping, Sweden	IAER
12	July 5, 1988	Toronto, Canada	MOEE
13	Jan. 31, 1989	Kansas City, KS	EPA Region 7 (Lab)
14	June 4, 1993	Bern, Switzerland	OFMET("Primary")
15	Oct. 20,1993	Dubendorf, Switz.	EMPA
16	Oct. 21, 1994	Ottawa, Canada	Env. Canada
17	Dec. 9, 1994	Prague, Czech Rep.	CHMI
18	Jan. 19,1996	Bern, Switzerland	OFMET ("Backup")
19	Nov. 20,1996	Braunschweig,Ger.	PTB
20	March 18, 1997*	London, England	NPL

*Since 1997, NIST has responded to requests for single SRPs from Australia, Spain, and, in 2002, 2 in Bureau Internationale des Poids et Mesures (BIPM) in France.

A visit to the NIST website on the SRP will provide details about the current system. It will also clarify that NIST has been working out an arrangement with the BIPM, with the goal of the BIPM taking over international support for the growing worldwide SRP network, over the next 5 years. NIST will still provide support to the U.S.(EPA) network.

EPA Traceability Protocol for ...Gaseous Calibration Standards-Status and Changes
Due to problems with the reliability/variability of the vendor-certified accuracy of the standard gases bought by state, local, and EPA Regional, and ORD laboratories for use to calibrate ambient air gaseous monitors, EPA established and has modified and expanded the scope of its Traceability Protocol for Certifying Gaseous Calibration Standards. In the late1980s, EPA ORD started reporting the results of a relatively small Protocol Gas verification program. Although sample size was small, probably not statistically representative, and had a relatively very low cost, vendors paid attention. This conclusion is indicated by the fact that the results improved over the 4-5 years of the program (paper at this meeting and session by John Schakenbach,U.S. EPA). Access to reports of the ORD verfication program can be found through the AMTIC website that contains the list of ORD reports and publications.

After ORD's QA service program divestment to OAQPS, the verification program was not continued. However, EPRI (ca.1998), and then EPA (2003-2004), in response to complaints by individuals from the user community, and some requests for re-institution of EPA-approved verification from some members of the gas vendor community, each performed an additional blind sampling study, and found that, without the program, significant problems, across pollutants, had again occurred. EPA is therefore looking into a vendor-funded, EPA-approved, 3rd party-operated, blind sampling, publicly-reported verification program.

Proposed Changes to 40 CFR Part 58- Brief Traceability Infrastructure Clarification

As part of the implementation of the current EPA Ambient Air Monitoring Strategy, currently under nationwide discussion, including CASAC subcommittee review, CFR changes have been proposed by OAQPS which include the addition of specific references, in 40 CFR Part 58, Appendix A, to the three components of the traceability infrastructure addressed in this paper. Updates and additions to the material currently in the websites are accessible at the following urls: http://www.epa.gov/ttn/amtic/qa or http://www.epa.gov/ttn/emc/news.html, for NPAP and the EPA Gas Protocol, respectively; and currently at the NIST website for the SRP, at http://www.cstl.nist.gov/nist839.03/ozone.html. Information about the EPA network can be obtained through the EPA author and network operating staff.

2003 EPA Audit of Protocol Cylinder Gases Abstract

John Schakenbach, U.S. EPA

Section 2.1.10 of "EPA Traceability Protocol for Assay and Certification of Gaseous Calibration Standards"(Protocol Procedures), September 1997 (EPA-600/R-97/121) states that EPA will periodically assess the accuracy of calibration gases and publish the results. Calibration gases are used to calibrate EPA reference methods which, in turn, are used to calibrate continuous emission monitoring systems (CEMs) that are used by regulated sources to report emissions to EPA. If the reference method calibrations are low by 10%, then emissions may be underreported by 10%. Therefore, accurate calibrations gases are critical in helping to ensure that required emission reductions are achieved.

Between 1978 and 1996, EPA's Office of Research and Development conducted several performance audits of calibration gases from various manufacturers. These audits had two goals, to provide a quality check for gas vendors and to connect users with good gas vendors. No EPA audits of calibration gases were performed between 1996 and 2002. One notable result in the most recent five years of audits is a steady, significant reduction in failure rate of the calibration gases, from about 27% in 1992 down to 5% in 1996.

In 2003, EPA performed an audit of calibration gases using National Institute of Standards and Technology (NIST) standard reference materials (SRMs) and NIST-traceable reference materials (NTRMs). The Agency arranged for third party purchases of gas cylinders made according to the Protocol Procedures. The cylinders were purchased and shipped to a third party for analysis in a manner that prevented gas vendors from determining that the gas cylinders were being audited. Forty two (42) source-level, tri-blend, EPA protocol, calibration gas cylinders were analyzed. The gas cylinders contained 50 ppm SO_2, 50 ppm NO, and 5% CO_2; 500 ppm SO_2, 400 ppm NO, and 12% CO_2; or 1000 ppm SO_2, 900 ppm NO, and 18% CO_2 in a balance of nitrogen gas. Gas cylinder documentation was also checked.

We encountered problems with some of the analytical techniques, especially related to NO. Chemiluminescence (API analyzer) was initially used to analyze the NO component in each cylinder. However, there was interference with the NO readings caused by the CO_2 component (quenching effect) which forced us to throw out these numbers. We then analyzed the NO using a UV (AMETEK) analyzer. However, there was some interference caused by the SO_2 in each cylinder. Based on interference testing, AMETEK developed a correction that we applied to the NO results. We then used FTIR (Spectral Insights, LLC) to determine NO, SO_2, and CO_2 concentrations for the cylinders that were off by 2% or more from the tag value. However, the corrected UV NO results were significantly different than the FTIR results, leading us to further question both sets of results. The SO_2 and CO_2 analyses (both using NDIR and FTIR) are in better agreement with each other, and easier to interpret. Results are pending, but will hopefully be available by the conference date. Results will be accompanied by some discussion of measurement methods and guidance on interpretation. A possible outline for an ongoing audit program is discussed.